MASTER EN SCIENCES INFORMATIQUES

Réalisation d'un proxyServeur d'anonymisation

Liénart, Stéphane

Award date:
2002

FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX, NAMUR

Institut d'Informatique
Année Académique 2001-2002

Réalisation d'un ProxyServeur d'Anonymisation

par ***Stéphane Liénart***

Mémoire présenté
en vue de l'obtention du grade de Licencié en
Informatique

Septembre 2002

Résumé

Les traitements invisibles des navigateurs permettent à des tiers de suivre et profiler, à son insu, un internaute particulier lors de ses navigations sur le Web. Pour se prémunir de ces indiscrétions vis-à-vis de ses propres informations personnelles, l'internaute peut, à juste titre, faire appel aux technologies de renforcement de la vie privée.

Nous proposons, dans ce mémoire, une étude de ces malicieux traitements invisibles et une analyse critique de leur portée sur la vie privée des citoyens du Web. Nous présentons également les diverses solutions qui s'offrent à eux aujourd'hui pour préserver leurs données à caractère personnel et ainsi améliorer la protection de leur vie privée.
Bien sûr, nous n'en resterons pas là puisque l'objectif final de ce travail est de réaliser un outil, en l'occurrence un proxyserveur d' « anonymisation », qui devrait permettre à ses utilisateurs de contrôler les traces qu'ils laissent sur le Web et dont la configuration par défaut sera la plus respectueuse pour leur vie privée.

Mots Clés : vie privée, Internet, traitements invisibles, navigateurs, scripting côté client, bavardage des navigateurs, cookies, Web bug, techniques de collecte de données personnelles, traitement des données à caractère personnel, TCP/IP, HTTP, HTML, cybermarketing, profilage, technologies de renforcement de la vie privée, P3P, confidentialité, cryptographie, infomédiaire, label Internet, serveur proxy, Java, Jigsaw, W3C

Remerciements

Je voudrais, avant tout, exprimer ma gratitude à Monsieur le Professeur Jean Ramaekers et Monsieur Jean-Marc Dinant qui ont accepté de diriger ce mémoire et qui m'ont apporté leur aide et leur expérience au cours de sa réalisation.

Je sais gré à Messieurs les membres du Jury pour leur lecture attentive et critique de ce travail.

Je souhaite aussi exprimer ma reconnaissance à ma mère pour ses conseils et sa lecture attentive du manuscrit.

Je remercie vivement mes camarades de promotion qui m'ont encouragé tout au long de l'année.

Merci à mes amis pour leur soutien.

Enfin, je tiens à remercier toute ma famille pour son soutien dans les moments difficiles, ses encouragements constants et sa patience.
Encore un grand merci à mes parents, qui m'ont permis de mener à bien ces deux années d'études en Informatique.

Mes dernières pensées iront à Soraya, je désire lui exprimer toute ma gratitude.

Table des Matières

Introduction

La simplicité de la navigation sur le réseau mondial et la liberté d'expression que l'on y trouve ne sont pas synonymes de respect et de protection de la vie privée (cfr. [3]). Pourtant, la plupart des internautes ignorent tout de ce qui se cache derrière ce monde virtuel dont «l'essentiel est invisible pour les yeux » (cfr [1]).

Surfer en toute simplicité sur les hyperliens de la toile ne protège pas les internautes des indiscrétions de certaines entreprises qui peuvent facilement profiter de la méconnaissance de ceux-ci pour soutirer, à leur insu, des informations les concernant, et d'en faire le commerce.

Que nous allions sur Internet dans le cadre de nos activités professionnelles ou depuis notre domicile, nos déplacements dans le monde d'Internet, si vaste et si riche, nous donnent sans doute une impression de liberté. Et le fait est que nous sommes libre d'aller n'importe où sur le réseau avec le sentiment de naviguer dans un parfait anonymat. L'absence d'intermédiaire visible peut renforcer ce sentiment d'absence de surveillance. La méconnaissance du fonctionnement des réseaux peut se trouver à l'origine de cette conviction, malheureusement fausse, que les connexions ne laissent pas de traces. Le préjugé très répandu qu'Internet est complexe peut donner l'assurance, hâtive, que si traces il y a, les moyens de les exploiter sont disproportionnés au regard de l'intérêt qu'on aurait à le faire. Mais nous verrons que la réalité est autre. Non, l'absence de traces n'est pas la règle sur Internet. Oui, il est possible d'être plus ou moins surveillé, peut-être pas étroitement, mais suffisamment pour que des outils de surveillance existent et que cette surveillance puisse être le fait d'acteurs de nature très variée... Il ne s'agit pas d'affirmer que si nous laissons des traces, la surveillance serait omniprésente. Si nous cessions juste de qualifier Internet de "virtuel", pour le considérer sous son angle bien réel, à savoir informatique et technique, nous constaterions qu'en matière de traces, Internet fonctionne exactement comme tout autre lieu du monde : ne pas laisser d' « empreintes » derrière nous demande des efforts, ce n'est que rarement garanti. La surveillance nécessite également des moyens, plus ou moins importants ; elle n'est pas inéluctable, mais elle est toujours possible.

De plus en plus, l'information que l'on trouve sur Internet n'est pas véritablement identique pour tout le monde, elle n'est plus statique, invariable. Ainsi, les pages HTML consultables sur le World Wide Web (ou Web) peuvent varier selon l'individu qui la consulte, en relation avec la connaissance qu'une société (de marketing, par exemple) possède déjà à son sujet. On en arrive dès lors à pouvoir transmettre des informations différenciées à chaque internaute explorant un site, les pages visitées deviennent personnalisées et les publicités qui lui sont présentées vont différer selon son profil de consommateur. C'est le but avoué du marketing One-to-One, faire de la publicité « à la carte », ciblée, spécifique pour chacun, selon ses goûts et ses intérêts (chose qu'aucun autre média n'avait offert jusqu'alors avec une telle facilité et efficacité). Un autre avantage majeur de la publicité sur Internet est que les

coûts de distribution sont faibles et proportionnels à l'efficacité[1]. Par conséquent, ces sociétés n'hésitent pas à employer certaines technologies pouvant nuire au respect de la vie privée des internautes visitant les nombreux sites supports sur lesquels apparaissent leurs bannières de publicité (sites qui peuvent être organisés en réseau, les « *ad networks* ») pour obtenir un maximum de renseignements les concernant.

Sur Internet, toute page HTML consultée et toute action posée par un internaute laisse une trace, une information, que des agences de commercialisation d'espaces publicitaires peuvent facilement recueillir afin de l'exploiter dans un but de marketing direct. Envoyer du courrier électronique (ou e-mail), taper des mots-clés sur un moteur de recherche, discuter sur un chat, ou poster un message dans un forum, accéder à un site Web, télécharger un fichier quelconque, cliquer sur une bannière publicitaire, tout acte exécuté sur Internet génère une information qui sera stockée dans la mémoire d'un ou plusieurs serveurs ayant été les acteurs de la communication, même si ceux-ci ne sont pas directement en contact avec l'internaute, ou pis dans la mémoire de l'ordinateur de l'internaute lui-même. Il ne reste plus qu'à rassembler ces informations, à les comparer et les analyser pour établir un profil très précis et avéré de son comportement, de ses habitudes de consommation, de ses goûts, de ses centres d'intérêt et d'autres données à caractère personnel concernant n'importe quel internaute particulier. Ces informations pourront ensuite être vendues au plus offrant ou échangées avec d'autres firmes de cybermarketing. Ainsi, il est devenu aisé pour ces firmes de profiler plus de 100 millions de personnes et de connaître pour chacune de ses personnes des renseignements tels que, par exemple : son âge, son sexe, le pays où elle habite, ses marques préférées, ses goûts musicaux, son orientation sexuelle, ses habitudes de consommation, son adresse e-mail, et dans certains cas, sa religion, ses loisirs, certaines de ses opinions, son éducation et ses études, ses derniers employeurs, son titre et sa fonction dans l'entreprise où elle travaille, le secteur d'activité, la taille et le chiffre d'affaires de cette entreprise,...
Il n'y a qu'à imaginer les informations que ces sociétés obtiendraient lorsqu'un internaute souhaite consulter le site d'un syndicat auquel il désire adhérer, d'un parti politique dont il se sent proche, d'un centre médical sur une maladie invalidante dont il est atteint, d'une religion, d'un groupe de pensée...

La collecte de données sur Internet peut être directe ou indirecte ; en d'autres termes, elle peut être effectuée soit à l'occasion d'un contact avec un interlocuteur, soit à l'insu de la personne concernée, souvent de manière automatique. La nature des données collectées varie selon le protocole utilisé sur le réseau, c'est à dire, selon le type de service. Dans la pratique, différents protocoles peuvent être utilisés ensemble pour accroître la rentabilité ou la qualité des échanges. Bien entendu, la façon la plus évidente de recueillir des éléments supplémentaires sur les utilisateurs est d'interroger directement ceux qui consultent effectivement le site. C'est ce qui se fait couramment avec les formulaires d'enregistrement en ligne.

Sans oublier que cette collecte d'informations est rendue possible par l'ensemble des processus qui permettent la communication avec les autres ordinateurs et qui laissent des traces invisibles pour l'utilisateur. Or, ces divers processus s'exécutent en toute discrétion, sans que l'utilisateur en soit informé et sans qu'il puisse avoir un réel contrôle sur les données échangées. C'est comme cela que l'on aboutit petit à petit à ce qu'on appelle « l'inversion du paradigme client/serveur » (cfr. [1]).

---

[1] La publicité étant facturée au nombre de contacts, c'est à dire au nombre exact de fois où la bannière aura été affichée sur l'écran d'un utilisateur. De plus, l'annonceur est assuré que son message ne sera en principe délivré qu'aux personnes correspondant précisément à la cible qu'il recherche.

Mais quelles sont ces traces que l'internaute « sème » sur son passage lors de sa navigation et comment permettent-elles la connaissance de tant d'informations personnelles concernant le surfeur ? Où se trouvent la réalité et le mensonge en ce qui concerne la communication de données relatives aux internautes à leur insu ? Nous verrons dans le premier chapitre de ce mémoire qu'un site peut collecter des informations sur un internaute de différentes manières. Lorsqu'il demande l'affichage d'une page d'un site, une certaine quantité d'informations est incluse automatiquement dans la demande transmise par le navigateur en son nom. Une fois la page reçue, le site peut demander des informations supplémentaires au navigateur. Pendant que la page est transmise, le site peut suivre sa trace en prenant des notes sur son comportement et en les stockant dans une zone de son disque dur (cookies) qu'il peut lire ultérieurement. Lorsque l'internaute remplit et renvoie un formulaire, les informations qu'il contient parviennent au site. Nous décrirons chacun de ces aspects en détail. Ainsi, nous allons exposer les différentes techniques de collectage de données à caractère personnel sur Internet, celles-ci n'ont d'ailleurs aucun équivalent dans le monde réel. Nous verrons par la même occasion en quoi ces techniques constituent une menace de plus en plus grande pour la protection de la vie privée des utilisateurs. De plus, nous constaterons qu'aucune des techniques mises en oeuvre n'est particulièrement lourde ni ne nécessite d'investissements colossaux. Bien au contraire ! Ensuite, dans un deuxième chapitre, nous ferons un large tour d'horizon et un bilan des différentes solutions pour se prémunir de ces « indiscrétions ». Bien sûr, nous en profiterons pour faire une analyse des véhicules d'Internet que sont les navigateurs (ou « browsers ») en les comparant sous le regard de la protection de la vie privée (cfr. [7] pour une « définition » du concept général de vie privée tel qu'il sera utilisé tout au long de ce mémoire). Pour conclure, nous exposerons dans un troisième et dernier chapitre les objectifs et les choix fixés par ce mémoire ainsi que leurs motivations.

Chapitre I : Des traces sur Internet... Les techniques de collecte de données à caractère personnel sur Internet

Dans ce chapitre, nous décrirons les « traitements invisibles » (à savoir les opérations accomplies par un programme à l'insu de son utilisateur) effectués par un navigateur hébergé sur l'ordinateur d'un internaute[2]. Nous considérerons plus particulièrement le cas d'un utilisateur ordinaire qui n'a pas modifié la configuration par défaut des paramètres de sécurité de son navigateur (cas représentant la majorité des internautes). Notre attention se portera sur les protocoles TCP/IP (indispensables pour communiquer sur un réseau du type d'Internet) et le protocole HTTP (protocole applicatif utilisé par les navigateurs et les serveurs) du point de vue client. Nous n'aborderons que succinctement certains des autres services disponibles sur Internet, citons à titre indicatif les forums, le chat, les newsgroups, le courrier électronique, les jeux « on-line », le e-commerce,... Ces derniers ont en effet plusieurs similarités avec celui que nous allons étudier plus en détail, à savoir la navigation sur le World Wide Web (cfr. [30] pour de plus amples informations sur le Web et ses standards), service le plus utilisé, et le plus connu, sur Internet. En fait, HTTP est un protocole suffisamment représentatif des autres en ce qui nous concerne. Néanmoins, il est possible que nous évoquions succinctement les risques d'atteinte à la vie privée des internautes que peuvent constituer les spywares, les virus et autres trojans, le spamming, ou encore les nombreuses autres menaces potentielles susceptibles d'affecter un utilisateur se connectant sur Internet.

Nous aborderons tout d'abord les traitements invisibles suivants, un par un : les protocoles TCP/IP, le bavardage des navigateurs, les cookies, le scripting côté client, la redirection automatique, les hyperliens invisibles et la mémoire cache. Ensuite, pour mieux se rendre compte de leur efficacité et de leur facilité d'utilisation, nous étudierons un cas concret de profilage d'un internaute les utilisant (presque) tous ensemble. En effet, lorsqu'ils sont pris chacun individuellement, la plupart des traitements invisibles (cfr. [1, 2, 5, 6, 8, 10, 11]) que nous évoquerons dans ce chapitre ne constituent relativement pas un grand danger vis-à-vis de la vie privée de l'internaute. Par contre, lorsqu'ils sont utilisés ensemble, ils peuvent devenir des éléments indispensables d'un mécanisme de profilage extrêmement efficace utilisé systématiquement et journellement sur plusieurs millions d'internautes par les entreprises de cyber-marketing.

Constatons aussi que ces techniques n'ont pas été conçues à l'origine pour nuire aux internautes ou porter atteinte à leur vie privée, mais qu'elles ont chacune leur justification et leur utilité. De plus, elles ont chacune leurs propres limites, c'est pourquoi elles ne se révèlent véritablement « dangereuses » vis-à-vis de la protection de la vie privée que lorsqu'elles sont utilisées conjointement. Nous indiquerons quelle attitude il paraît approprier d'adopter lorsque l'on fait d'Internet un outil de travail, de divertissement ou d'information au quotidien. En effet, l'expérience montre que l'on est beaucoup plus conscient d'un risque si l'on en comprend les tenants et les aboutissants.

Sur les autoroutes de l'information, l'enjeu, c'est la protection de notre vie privée. La solution principale, c'est de connaître les risques et d'agir en conséquence. En un mot, d'être vigilant !

---

[2] Nous ne prétendons pas être exhaustif étant donné les nouvelles techniques apparaissant régulièrement dans ce domaine.

I.1. Les protocoles TCP/IP

L'internaute moyen ne connaît pas ou peu de choses concernant les protocoles qui contrôlent le réseau qu'il utilise. Tout au plus a-t-il entendu parler d'adresse IP.

La communication entre ordinateurs sur le réseau mondial n'a été rendue possible que grâce au développement et à l'utilisation de protocoles standardisés. Les protocoles TCP/IP (Transmission Control Protocol/Internet Protocol) sont sans aucun doute les plus importants d'entre eux et sont au coeur de l'ensemble des services offerts par les applications permettant d'utiliser Internet, car tous ces services sont tous basés sur TCP/IP pour la transmission des données[3]. En effet, ils permettent le transport des petits paquets d'informations, envoyés sous forme d'octets, que deux ordinateurs s'échangent au travers d'un vaste réseau hétérogène d'ordinateurs.

I.1.1. Adresses IP et Domain Name Service

Cette communication est rendue possible grâce aux adresses IP qui identifient chaque ordinateur individuellement sur l'entièreté d'un réseau TCP/IP (rappelons qu'elle détermine univoquement un ordinateur sur le réseau au moins pendant toute la durée d'une session). Chaque paquet échangé comporte l'adresse IP de l'expéditeur et du récepteur (les interlocuteurs sont toujours informés de l'adresse IP de l'autre partie car il en a besoin pour savoir où envoyer les informations demandées). Ainsi, quoiqu'on fasse sur le réseau Internet, tous les paquets que notre ordinateur envoie et qu'il reçoit contiennent son adresse IP et celle du serveur avec lequel il établit une connexion ou est déjà connecté. En ce qui nous concerne, cette adresse IP peut être attribuée à notre machine de différentes manières, selon le type de notre accès à Internet. Si nous nous connectons dans une entreprise ou dans une organisation plus importante (comme une université ou une administration publique), nous avons alors vraisemblablement une adresse IP définitive (ou permanente). On parle alors d'adresse IP fixe, ou statique. Un serveur Web a nécessairement une adresse IP fixe. Si nous sommes un particulier se connectant de chez lui, c'est probablement une adresse différente qui nous sera attribuée automatiquement et aléatoirement par notre fournisseur d'accès à chaque nouvelle connexion et pour la durée de la session. On dit alors que notre adresse IP est dynamique[4] (car elle est temporaire). Un site visité peut donc déterminer qu'un membre de tel ISP vient de demander une de ses pages mais ignore lequel. Signalons que les adresses IP sont totalement distinctes et sans aucun rapport avec les adresses e-mail. Notre adresse e-mail est l'adresse à laquelle est envoyé le courrier électronique nous étant destiné, elle nous identifie de manière unique (il ne faudrait pas en effet qu'un e-mail confidentiel dont nous sommes le destinataire soit reçu par d'autres personnes sans que ce soit intentionnel). Notre adresse IP, par ailleurs, est une adresse temporaire (dans la plupart des cas) et est utilisée pendant la durée de la session pour obtenir les pages demandées. Elle est un peu l'analogue du numéro de téléphone d'une cabine publique utilisée pour passer un appel.

---

[3] Ces protocoles ont été conçus pour être simple à utiliser, et sont indépendants de tout type d'ordinateur et de tout système d'exploitation.

[4] Parfois des adresses IP statiques font référence au même utilisateur pendant longtemps. Les adresses IP statiques sont souvent employées quand on utilise des technologies alternatives (ADSL, câble TV) du fait que la connexion est permanente. Comme ces technologies deviennent de plus en plus communes, l'emploi des adresses IP statiques devient relativement plus fréquent.

Nous ne détaillerons pas davantage les principes du fonctionnement des protocoles TCP/IP qui régissent les communications sur l'ensemble du réseau Internet (ce n'est d'ailleurs pas le but de cette section). Nous souhaitons juste constater l'existence de quelques mécanismes sous-jacents qui sont souvent méconnus du grand public. C'est pour cette raison, entre autre, que, pour un utilisateur néophyte, l'essentiel sort de son ordinateur bien souvent sans qu'il en soit conscient.

Le nombre d'adresses IP est actuellement limité. C'est par le biais d'une procédure internationale[5] que les adresses IP sont attribuées dans le monde aux fournisseurs d'accès Internet, qui les réassignent à leurs clients, organisations ou particuliers. Par l'emploi d'un outil de recherche disponible à tous, tel que par exemple http://www.ripe.net/cgi-bin/whois, il est possible d'identifier le responsable de l'attribution d'une adresse IP spécifique. En général, ce responsable sera :

- Le responsable d'un réseau local relié à Internet, comme une entreprise ou une administration publique. Dans ce cas, elle utilisera probablement un système d'adressage IP fixe et conservera une liste de correspondance entre les ordinateurs individuels et les adresses IP. Mais cette liste peut aussi être tenue à jour automatiquement si ce responsable emploie un serveur DHCP[6] (Dynamic Host Configuration Protocol).
- Un fournisseur d'accès à Internet (ISP Internet Service Provider) qui a un contrat avec un utilisateur d'Internet. Dans ce cas, le fournisseur d'accès Internet tiendra en général un fichier à jour contenant les adresses IP attribuées, l'identité de l'utilisateur, la date, l'heure, et la durée de l'attribution de l'adresse. En outre, si cet utilisateur se sert d'un réseau publique de télécommunication (téléphone mobile ou non), le numéro appelé (et les date, heure et durée) sera enregistré par la société de téléphonie dans un but de facturation.
- Le dépositaire du nom de domaine (voir ci-dessous), qui peut être le nom d'une entreprise, le nom d'un employé d'une entreprise, ou d'un individu.

Dans ces cas, on voit qu'avec l'assistance du tiers responsable de l'attribution, un internaute (son identité civile : nom, adresse, numéro de téléphone) peut être identifié par des moyens raisonnables.

Beaucoup d'utilisateurs sont tentés de croire que les informations envoyées d'un ordinateur à un autre (par exemple, leur courrier électronique) suivent le chemin le plus direct. Or le lien entre le site que le surfeur consulte et sa propre machine est complexe. Les données adressées à une machine ont souvent un long chemin à parcourir avant d'arriver à destination, chaque étape étant un routeur[7] qui aura souvent (au moins) un droit de regard sur l'entête du

[5] La Société Internet pour les noms et nombres alloués (Internet Corporation for Assigned Names and Numbers ICANN) est l'organisation sans but lucratif formée pour être responsable de l'allocation d'espace aux adresses IP (http://www.icann.org). En Europe l'espace d'adressage est géré par le RIPE (Réseaux IP européens) (http://www.ripe.net).

[6] Le protocole de configuration d'hôte dynamique est un protocole Internet qui automatise la configuration des ordinateurs utilisant TCP/IP. DHCP peut être employé pour assigner automatiquement des adresses IP (http://www.dhcp.org).

[7] Un routeur est un mécanisme important qui fournit des routes pour les réseaux TCP/IP. Ceci signifie que la route du TCP/IP est dynamique, et peut changer en fonction d'une surcharge ou d'un échec rencontrés sur certains routeurs ou liens. Il peut garantir spécialement que seules des adresses IP autorisées peuvent émaner

paquet acheminé (néanmoins, on peut, la plupart du temps, les considérer comme des éléments neutres, bien que des moyens particuliers puissent être utilisés pour y intercepter le trafic Internet). Bien entendu, il en va de même pour les e-mails ou tout autre type de trafic sur le réseau. Mais ce n'est pas tout, la technologie n'est pas le seul intervenant dans le chemin que suivent les paquets TCP/IP. Les différents « territoires » des ISP n'autorisent le passage que de certaines données[8]. Donc plusieurs opérateurs, appartenant à des pays différents, vont intervenir dans l'acheminement des données et ces dernières non seulement ne suivent pas un trajet direct mais encore plusieurs paquets d'informations constituant le même courrier électronique peuvent suivre, dans certains cas de figure, des chemins complètement différents. Chacun de ses acteurs participant à l'échange d'informations pourraient éventuellement associer l'adresse e-mail de l'utilisateur avec l'adresse IP de sa machine, ou conserver une copie de l'entièreté (ou, plus rarement, d'un morceau[9]) du message. Ainsi, un courrier électronique classique s'apparente plus à une carte postale qu'à une lettre cachetée.

L'adresse IP, qui est un quadruplet de chiffres compris entre 0 et 255, n'est pas aisée à retenir. Pour faciliter la navigation, on a inventé les noms de domaine[10] et les services associés (DNS). Un nom de domaine n'est pas en soi nécessaire pour pouvoir connecter un ordinateur à Internet. Ces derniers ont pour mission d'obtenir l'adresse IP équivalente au nom de domaine tapé par l'utilisateur sans que celui-ci doive intervenir. Ainsi lorsque celui-ci tape une URL (Uniform Ressource Locator, càd l'adresse d'une ressource Web) dans son navigateur, il pense accéder directement au site Web correspondant. Or il ignore qu'avant cela, une série d'opérations ont été effectuées par son ordinateur sans qu'il n'en sache rien. Parmi ces opérations, sa machine a interrogé un serveur distant, un serveur DNS (généralement il s'agira du serveur DNS du fournisseur d'accès de l'utilisateur), serveur dont elle connaît au préalable l'adresse IP[11]. Ce dernier lui a ensuite répondu en lui fournissant

---

d'un certain fournisseur d'accès Internet. Il est important de noter que pour le routage au travers de réseaux TCP/IP, la vitesse de transmission est le critère le plus valable. Dans la mesure où les informations circulent sur des câbles à une vitesse presque égale à celle de la lumière, il peut être plus efficace d'acheminer les paquets TCP/IP de Londres à Madrid en passant par New York s'il y a un embouteillage dans le réseau à Paris. La route suivie suit donc une logique de performance et, en matière de télécommunication, la performance est liée bien plus à la congestion du réseau qu'à la distance séparant les nœuds de télécommunication (les routeurs). L'internaute n'a aucun moyen de modifier cette route, même s'il peut savoir quelle route est suivie entre deux points sur le réseau à un certain moment avec certains outils, comme traceroute.

[8] Ainsi, les accords entre certains fournisseurs d'accès vont amener, par exemple, l'e-mail de X destiné à Y à faire un détour par les Etats-Unis, l'Angleterre puis la France alors que les deux interlocuteurs résident en Belgique. Ces accords font que le réseau Internet est divisé en zones souvent « politiques », qui empêchent la circulation libre de certaines informations en leur sein. Ceci explique pourquoi il ne faut donc pas s'étonner de voir un courrier électronique envoyé en Belgique faire un petit détour par trois pays étrangers avant d'arriver chez son destinataire.

[9] En théorie, la route suivie par les paquets TCP/IP est dynamique et peut donc changer au cours de l'importation d'une page Web ou de la transmission d'un courrier, mais en pratique elle reste généralement statique.

[10] Ces noms sont constitués de trois parties : un nom, un point et une désinence. Le nom est une chaîne de caractères, qui peut être formée d'une ou de plusieurs sous-chaînes séparées par des points ; la désinence peut être un domaine générique tel que "com" pour les sites Web commerciaux, "org" pour des organisations sans but lucratif, ou un domaine géographique comme "be" pour la Belgique. Le nom de domaine est payant, et les sociétés ou individus qui souhaitent en avoir un doivent s'identifier au préalable.

[11] C'est l'un des paramètres de l'ordinateur que l'utilisateur a dû configurer lorsqu'il a pris un abonnement auprès de son ISP, à moins que ce paramétrage n'ait été fait automatiquement par un programme spécifique (diffusé sur CD-ROM), comme le font désormais de nombreux fournisseurs d'accès grand public, ou encore par l'informaticien qui administre le réseau de l'entreprise où l'internaute se connecte.

l'adresse IP du site associé au nom de domaine contenu dans l'URL tapée précédemment[12] (c'est ce qu'on appelle effectuer la "résolution" (conversion) des noms de domaine en adresses IP). Par conséquent, on peut très bien imaginer que les personnes responsables du serveur DNS peuvent tout savoir des sites visités par notre utilisateur. Il est aussi possible que l'association adresse IP/nom de domaine soit erronée (pourquoi pas volontairement ?). Dans ce cas, l'internaute risque de voir un autre site que celui qu'il voulait consulter. Par conséquent, les internautes doivent être prudents en supposant la validité de l'association entre une adresse IP et un nom DNS.

I.1.2. Traçage des adresses IP par audit

Considérons la situation suivante (typique d'une navigation courante sur Internet) : nous cliquons sur un lien. Notre navigateur envoie au site Web une requête HTTP qui sera divisée en plusieurs paquets IP, chacun d'entre eux contient notre adresse IP en tant qu'émetteur et celle du serveur en tant que récepteur. Le contenu de la requête indique au serveur en question la page sur laquelle nous désirons nous rendre. Certes, le serveur va servir cette requête et nous envoyer sa réponse (à savoir la page Web que nous avons demandé et qui s'affichera alors sur notre écran), mais rien ne l'empêche, dans le même temps, d'enregistrer dans un fichier d'audit, le moment où il a reçu notre requête, l'adresse IP de l'ordinateur qui a fait la requête (à savoir le nôtre), et quel fichier a été envoyé (la page Web demandée). De plus, si personne ne nous l'indique par ailleurs, nous ne saurons jamais qu'un tel enregistrement a été fait. Après cela, traiter et exploiter ces enregistrements n'est plus qu'une question d'organisation et d'utilité. Pourquoi ne pas rassembler ces informations, les classer chronologiquement et les analyser pour reconstruire notre parcours sur le site (pour établir un profil de nos centres d'intérêts, par exemple, ou pour des raisons purement statistiques). Pour établir ce parcours, rien de plus facile : il suffit de récupérer notre adresse IP via la variable d'environnement que le navigateur envoie et de rechercher dans le fichier d'audit existant les entrées qui la comportent.

Un enregistrement du fichier d'audit peut contenir les informations suivantes (contenues dans la requête d'un client et dans la réponse du serveur) : l'adresse IP du client bien entendu mais aussi le moment de la connexion, la méthode utilisée pour la requête HTTP ("GET","POST", ...), le nom de l'objet demandé (page, image, fichier, etc.), la réponse que le serveur a fournie au client, le volume en octets des fichiers transférés entre le serveur et le client, la page de provenance de l'utilisateur lorsqu'il a fait sa requête et enfin la signature du navigateur qui a effectué la requête.

Enfin, il faut savoir que le fichier constitué à l'aide de cette fonctionnalité de traçage des adresses IP peut être enrichi durant notre session si nous fournissons nous-mêmes des informations : si nous remplissons un formulaire, toutes les informations que nous envoyons peuvent être associées à notre adresse IP. Ainsi nous avons indiqué notre nom, notre adresse IP devient une étiquette parfaite. Il en est de même si nous avons envoyé un e-mail. La technique peut se révéler encore plus malicieuse lorsqu'elle est associée avec les fameux cookies (cf. infra).

Quelle peut être l'utilité de ces fichiers d'audit ? Collecter les requêtes reçues sur un serveur doit permettre, en principe, à l'administrateur du serveur de connaître avec précision

---

[12] Signalons que cette opération a lieu à chacune de nos propres requêtes. En résumé, notre navigateur lance une requête au serveur DNS, récupère l'adresse IP du serveur auquel nous demandons une page et contacte ce serveur.

la répartition des charges du système. Savoir, quand le serveur est le plus mis à contribution, connaître les fichiers les plus souvent téléchargés,... Il sera alors amené à concevoir différemment l'architecture et l'ergonomie de son site, pour raccourcir le temps de chargement des pages les plus demandées par exemple. L'audit a donc pour finalité d'optimiser le fonctionnement du site.

Mais une des caractéristiques de l'audit, c'est que l'internaute n'a absolument aucune prise sur lui. Il ne peut pas savoir que le serveur abritant le site qu'il consulte est en train d'enregistrer ses transactions. Le traçage en lui-même, tout comme son traitement et l'utilisation du fichier ainsi constitué peut être effectué entièrement à son insu. La question est donc de savoir dans quelles circonstances de tels fichiers sont habituellement utilisés. Il est compréhensible que l'administrateur du serveur tente d'améliorer le fonctionnement de son site en analysant les connexions effectuées et, dans ce cadre, il n'a aucune raison de conserver toutes ces traces, assez gourmandes en espace disque. Il détruira donc périodiquement les fichiers devenus inutiles ou trop âgés. Si une optimisation n'est pas prévue durant certaines périodes, il désactivera la fonctionnalité d'enregistrement, devenue superflue. Néanmoins, il gardera toujours la faculté d'observer (et d'enregistrer) le contenu des requêtes sur le serveur dont il s'occupe.

Bien sûr, un tel outil peut également être utilisé pour des fonctions autres que purement techniques. C'est grâce à lui que l'on peut mesurer l'audience d'un site, produire des statistiques précises quant aux pages visitées. L'atteinte à la vie privée ne se situe sans doute pas dans le fait de produire des statistiques mais dans la possibilité d'utiliser des informations, dans un but commercial par exemple, relatives aux comportements des internautes sans qu'ils en soient au courant ni qu'ils puissent s'y opposer.

De surcroît, les informations fournies par l'audit ne sont pas si anodines. Lorsque l'internaute utilise une adresse IP statique, celle-ci identifie de façon unique son ordinateur (et donc son utilisateur par la même occasion). Dans ce cas, tout ce qu'il fait sur le site peut être mémorisé et mis en relation d'une session à l'autre sur le serveur. Ces informations qui peuvent décrire ses intérêts ou d'autres aspects de sa personnalité peuvent être échangées, analysées, exploitées pour des raisons commerciales ou autres, cédées à des tiers,... Si maintenant l'adresse IP de l'utilisateur est dynamique (elle n'identifie pas l'internaute, mais elle identifie son fournisseur), il est toujours possible de retrouver le lien jusqu'à lui, en passant par le fournisseur d'accès. Si ce fournisseur conserve un journal des connexions, il saura qui a eu telle adresse IP à tel moment pendant telle durée. Néanmoins, les fournisseurs d'accès donnent rarement la possibilité d'examiner ces données, sauf lorsqu'ils sont contraints par la loi. Il y a donc peu de chances que l'on puisse obtenir des informations nominatives sur l'internaute au moyen de cette technique, à moins que ce dernier ne les fournisse de lui-même.

Si les informations récupérées ne sont pas forcément riches, elles peuvent tout de même être judicieusement utilisées pour peu que, lors de sa connexion au site, l'internaute a été amené à en fournir d'autres de son plein gré, en remplissant un formulaire comportant des informations plus précises à son sujet, par exemple. A ce moment là, le lien entre son adresse IP et les informations qu'il a introduite peut être fait aisément. Lorsqu'un utilisateur fourni une information personnelle, celle-ci peut aider à faire le lien avec l'ensemble de son parcours sur le site auquel il a procuré cette information et peut donc conduire à l'élaboration d'un profil précis. De plus, si le site en question a un contenu d'ordre religieux, médical ou politique, l'utilisateur n'a pas forcément envie que le serveur conserve la trace de son chemin, qui peut être très révélateur de ses attentes. D'où la nécessité pour les personnes visitant ce site d'être certaines que tout usage du fichier d'audit se situe dans un cadre légal : avoir un

droit d'opposition et un droit d'accès, être mis au courant de son existence et du but de son emploi.

Sur Internet, tout (FTP, chat, newsgroup, e-mail...) peut être tracé par le serveur. Mais ne sombrons pas dans la paranoïa, n'oublions pas qu'il s'agit d'une méthode qui n'est pas nécessairement utilisée. Et même si elle l'est, ce peut être que pour des raisons d'optimisation technique, sans conservation des informations ou tentative d'analyse ultérieure plus approfondie. Tout le monde n'a pas de mauvaises intentions. Par contre, il convient de rester vigilant et seule une information exacte, transparente et précise des internautes peut permettre de s'assurer que la vie privée est respectée.

I.1.3. Ping

Il existe aussi un programme, appelé Ping, qui permet d'envoyer des demandes pour savoir si une adresse IP particulière (quel que soit l'endroit sur le réseau où se situe l'ordinateur auquel est attribué cette adresse, s'il en existe un) est en cours d'utilisation. La machine « pinguée » répondra à n'importe laquelle de ses demandes provenant de l'extérieur sans que son propriétaire soit au courant de la provenance des demandes reçues. Celui-ci ignorera donc si quelqu'un, situé dans le réseau, a essayé de savoir s'il était connecté à un certain moment, et dans quel but. Un internaute malveillant peut même parvenir à crasher une machine connectée au réseau et utilisant un système d'exploitation relativement âgé rien qu'en effectuant cette commande ping[13] (avec en paramètre l'adresse IP de la machine en question et une taille de paquets supérieure à 64ko, taille normalement interdite).

Même si tous les processus de traitement des données exposés ci-dessus peuvent être légitimes, voire même, selon les circonstances, indispensables au bon fonctionnement du réseau Internet, l'utilisateur doit être au courant du fait que de tels processus ont lieu et que certaines mesures de sécurité sont disponibles.

Selon J.M. Dinant, « les informaticiens eux-mêmes mesurent rarement l'ampleur des effets de certaines de leurs décisions techniques sur la protection des données à caractère personnel » (cfr. [2]). En effet, du fait qu'Internet, depuis ses débuts, a été considéré comme un réseau ouvert, les protocoles de communication ont plusieurs caractéristiques qui peuvent mener à un manque de respect de la vie privée des internautes, et ceci moins à cause d'une décision délibérée, qu'à cause d'effets secondaires.

I.2. Le bavardage des navigateurs

Fondamentalement, les protocoles les plus utilisés sont :

- le HTTP (HyperText Transport Protocol) pour surfer sur Internet
- le FTP (File Transfer Protocol) pour transférer des fichiers
- le NNTP (News Network Transport Protocol) pour accéder à des forums d'utilisateurs, appelés aussi newsgroups.
- le SMTP (Simple Mail Transport Protocol) pour envoyer des e-mails.
- enfin, le POP3 ou l'IMAP, pour recevoir des courriers électroniques.

[13] Pour plus d'explications, voir The Ping Of Death : http://www.pp.asu.edu/support/ping-o-death.html.

Tableau I-1 : *Hiérarchie des protocoles utilisés sur Internet* (source : référence [2])

Couche Services	***Protocole et Usage***	HTTP Surfer	SMTP envoyer un e-mail	POP3 recevoir un e-mail	NNTP accéder aux news et en transférer	FTP importer ou exporter des fichiers	de nombreux protocoles actuellemer utilisés ou à venir
	Hardware	Client/Serveur HTTP	Client/serveur de courrier électronique		Client/serveur de news	Client/serveur FTP	
	Acteur	Utilisateur/fournisseur de service					
Couche TCP/IP	***Protocole***	TCP/IP					
	Hardware	Routeur					
	Acteur	Fournisseur d'accès Internet					
Couche matérielle	***Protocole et Usage***	PPP utilisé sur lignes téléphoniques analogiques	ADSL utilisé sur lignes téléphoniques actuelles	ETHERNET utilisé sur les réseaux locaux (LAN)	de nombreux protocoles actuellement utilisés ou en cours de développement		
	Hardware	Modem	Modem ADSL	Carte réseau Ethernet	...		
	Acteur	Opérateur de télécommunication					

Ces différents protocoles sont nécessaires du fait que le protocole TCP/IP permet seulement la transmission d'information en vrac d'un ordinateur vers un autre. L'ordinateur qui fournit un service est nommé un SERVEUR. L'ordinateur qui utilise un service est nommé CLIENT. Pour fournir un service technique, le client et le serveur emploient le même protocole, c'est-à-dire les mêmes règles de communication. C'est pour cette raison qu'Internet est un réseau client/serveur[14]. Il est important de souligner que quelque soit le service utilisé, le protocole TCP/IP est toujours employé, y compris pour tous les services mentionnés ci-dessus. Ce qui signifie aussi que toute menace pour la protection de la vie privée liée au protocole TCP/IP sera présente lors de l'emploi de n'importe quel service sur le Web. Cette section et les trois suivantes analysent des caractéristiques présentes sur presque toute mise en œuvre du protocole HTTP par les navigateurs les plus courants. Il faut noter que la combinaison de ces

[14] L'utilisation du réseau par les applications est basée sur le modèle client/serveur :
- L'application cliente (avec laquelle interagit l'utilisateur) aura besoin d'un service fourni par un serveur distant. Elle contacte ce dernier à travers le réseau pour lui soumettre une requête.
- Le serveur transmet les données demandées au client, par la même connexion réseau (celle-ci est bi-directionnelle).
- Une même machine peut être serveur pour plusieurs services simultanément. Ils sont distingués par le port que doit utiliser le client pour contacter tel ou tel service.
- Ces ports sont implémentés par l'utilisation d'entiers (entre 1 et 65536) auxquels on a souvent donnés un nom : par exemple, on parle sans distinction du "port 80", du "port HTTP" ou du "port Web" pour dénoter le port d'un serveur Web que doit contacter un browser (application cliente) pour obtenir la page HTML qu'on lui demander.

protocole HTTP par les navigateurs les plus courants. Il faut noter que la combinaison de ces caractéristiques peut avoir des conséquences sérieuses sur le respect de la vie privée des internautes. En effet, HTTP a une grande importance stratégique du fait qu'il s'agit du principal protocole employé sur le Web.

Ainsi, les données stockées sur les serveurs, sous forme de pages écrites en langage HTML sont transmises d'un ordinateur à l'autre à l'aide du protocole HTTP. Ce protocole permet d'échanger des messages entre le client et le serveur à l'aide de requêtes et de réponses HTTP (fonctionnement selon un mode requête-réponse). Les requêtes et réponses HTTP contiennent des en-têtes permettant d'envoyer des informations particulières de façon bilatérale. Or, comme nous le verrons dans cette section, certaines d'entre elles sont superflues vis-à-vis de la communication proprement dite.

Pour surfer sur Internet, tout utilisateur doit utiliser un programme de navigation, ce qu'on appelle un navigateur (càd un logiciel applicatif client HTTP qui permet d'entrer en relation avec un serveur HTTP), qui permet d'afficher « en clair » les pages implémentées en HTML. Ce sont ces navigateurs qui utilisent le protocole HTTP pour demander et télécharger les pages Web. Mais ce sont aussi ces derniers qui envoient et stockent des informations personnelles concernant l'internaute (et sa machine).

Pour qu'un navigateur puisse rechercher une page Web sur le réseau, il doit émettre une requête HTTP (*commande* GET + *l'adresse de la page*). Le contenu de cette requête est invisible pour l'utilisateur. Au niveau HTTP, chacune des requêtes émises par un browser contient un certain nombre d'informations, à commencer par ce que l'utilisateur a demandé : l'URL d'une page Web le plus souvent. Mais elle contient également de nombreuses informations sur l'environnement logiciel de l'ordinateur de l'internaute. Ces informations sont systématiquement transmises dans chaque paquet qui part de son ordinateur vers le serveur. Lorsque le paquet arrive sur le serveur, un « processus système » est créé sur la machine, processus qui intègre ces informations sous forme de variables d'environnement et exécute la requête demandée (ce qui aura normalement pour effet d'envoyer le code HTML de la page demandée, si elle est disponible). En résumé et de façon très simplifiée, lorsque le navigateur veut afficher une page, il accomplit les tâches suivantes :

- Il ouvre une connexion réseau (en pratique il ouvre un socket),
- Il envoie une requête HTTP, contenant la commande GET pour obtenir la page souhaitée ainsi que d'autres informations liées à l'environnement logiciel de l'internaute (voir ci-après),
- Il reçoit les données que lui envoie le serveur (à savoir la page Web demandée par exemple, mais aussi d'autres informations non sollicitées, comme un cookie (cfr. infra) par exemple),
- Il ferme la connexion.

Ainsi à chaque page qu'il veut afficher, il est obligé d'effectuer cette même séquence.

Pour visualiser le contenu de l'entête de nos requêtes HTTP et nous faire une idée de ce qu'elles peuvent « raconter » au sujet de notre environnement logiciel, nous n'avons qu'à consulter, par exemple, un des sites suivants :

http://moneycentral.msn.com/investor/test/user/httphead.asp
http://www.multiproxy.org/env_check.htm - http://www.anonymat.org/vostraces/
http://www.leader.ru/secure/who.html - http://privacy.net/analyze/

Les pages de ces sites s'adaptent en temps réel en fonction de l'information transmise par notre navigateur, information spécifique à l'environnement logiciel de notre ordinateur. Le contenu de ces pages n'est donc pas statique puisqu'il sera différent pour chaque ordinateur qui se connecte à l'un d'entre eux, on parle alors de « site polymorphique ».

Les paramètres se trouvant au sein de l'entête des requêtes HTTP méritent que l'on s'y intéresse de plus près car ils sont révélateurs du bavardage des navigateurs. Bien que ces normes contiennent des avertissements et des recommandations pour que leur utilisation respecte les données à caractère personnel de l'internaute, c'est dans les mains des concepteurs de logiciels de navigation que se trouvent la liberté de les appliquer ou non.

- HTTP_REFERER : ce paramètre contient l'adresse complète de la page précédemment consultée par l'utilisateur avant qu'il clique sur l'hyperlien menant à la page actuelle[15] (et ce quelque soit le type de cet hyperlien, visible ou invisible, cfr. section I.6) ; il s'agit donc de l'adresse de la page parente à la page visitée. En d'autres mots, ce champ permet au client de spécifier l'URL de la ressource à partir de laquelle l'URL de la requête en cours a été obtenue (à savoir les informations suivantes, sur la dernière page vue par le navigateur client : protocole utilisé, adresse DNS ou IP du serveur, chemin d'accès à la ressource, et ressource demandée (page, image, cgi ...)). Un site visité étant informé de la page consultée antérieurement, cela lui permet de connaître l'URL de la page du site qui a dirigé l'internaute vers lui ou, s'il s'agit d'une page du même site, ce dernier peut déterminer quelle section a été vue préalablement.
  Ce champ est vide si l'utilisateur a tapé lui-même l'URL de la page qu'il souhaite visiter dans la barre d'adresses de son navigateur.
- HTTP_USER_AGENT : il s'agit de la signature du navigateur. Il donne les caractéristiques du browser (nom du navigateur, son type, sa marque et son numéro de version (et de sous-version) ainsi que sa version linguistique). Il indique aussi la marque, le type et la version du système d'exploitation fonctionnant sur l'ordinateur de l'internaute.
- HTTP_ACCEPT : il spécifie certains types de fichiers qui sont acceptables comme réponse. Plus techniquement, il s'agit des différents types MIME supportés par le client HTTP. De cette façon, il communique le nom de certaines applications installées sur la machine de l'utilisateur.
- HTTP_ACCEPT_LANGUAGE : il indique les langues préférées par l'internaute dans la réponse à la requête (il s'agira la plupart du temps des langues parlées par le surfeur).
- HTTP_FROM : ce champ indique l'adresse e-mail de l'internaute utilisant le navigateur. Heureusement, la plupart des browsers actuels ne transmettent pas ce champ afin de préserver la vie privée de leurs utilisateurs, suivant ainsi les recommandations des auteurs du protocole HTTP.

La définition technique de ces champs peut être trouvée dans les RFC 2068 et RFC 2616 décrivant le protocole HTTP 1.1 (cfr. [21] et [23]). Signalons toutefois que ces différents champs sont facultatifs (seule la commande GET est vraiment indispensable), ils ne sont donc pas essentiels à la session HTTP entre les deux ordinateurs !

[15] Dans le cas d'une page fournissant les résultats d'une recherche sur un moteur de recherche, le nom de la page référente contient les critères de recherche (à savoir les mots-clés tapés par l'internaute). Dans le cas d'un hyperlien invisible, il s'agira de la référence de la page où sera affichée l'image, une bannière publicitaire par exemple.

Dans ces requêtes, le navigateur communique également l'adresse IP de la machine de son utilisateur (ou REMOTE_ADDR, elle est requise par l'utilisation des protocoles TCP/IP et est nécessaire pour que le serveur interlocuteur sache à qui il doit répondre). Un serveur recevant une requête HTTP d'un internaute peut aussi obtenir indirectement le nom de son fournisseur d'accès (contenu dans le nom DNS de la machine cliente, ou REMOTE_HOST, s'il est défini).

Enfin, il existe deux champs particuliers, appelés HTTP_SET_COOKIE (contenu dans une réponse HTTP) et HTTP_COOKIE (contenu dans une requête HTTP), dont l'usage est le suivant : lorsqu'un navigateur reçoit une réponse à une requête provenant d'un certain site, il stocke la valeur du champ HTTP_SET_COOKIE sur le disque dur de l'internaute (cette procédure est automatique). Lors de la prochaine visite de l'internaute sur le site en question, la valeur du cookie précédemment sauvegardé lui sera systématiquement retransmise dans chaque requête qui lui est destinée, au moyen du champ HTTP_COOKIE[16]. Le site pourra alors consulter cette valeur et la modifier (ou la supprimer).

Remarquons encore qu'un serveur Web peut également avoir accès, au moyen de scripts particuliers, à des informations telles que l'horloge locale, la résolution du moniteur, le nombre de processeurs et leurs modèles respectifs, la configuration du navigateur (s'il accepte les JavaScripts, les VBScripts et/ou les applets Java, la liste des plugins installés, et plus encore), le nombre de pages visitées durant la session courante du navigateur... Certaines de ces informations sont sensées aider les sites à présenter les pages demandées de manière optimale pour l'écran de l'internaute.

Néanmoins, il est important de signaler que ces paramètres d'environnement ont aussi une utilité (autre que celle, malicieuse, d'indiquer des données à caractère personnel à notre insu). Ainsi, ces informations sont transmises au serveur Web pour lui permettre de prendre en compte des éléments propres à la configuration de la machine d'un internaute. Connaître, par exemple, le type de navigateur que cet internaute utilise (et sa version) peut lui permettre de ne pas lancer certains scripts qu'il sait ne pas être compatibles avec ce navigateur. La variable contenant la référence de la dernière page visitée permet, par exemple, aux sites qui achètent des bandeaux publicitaires de comptabiliser le nombre de connexions qui ont été effectuées immédiatement après un clic sur une des pages comportant un tel bandeau, ce qui permet d'en évaluer l'efficacité.

Faut-il avoir peur de ces paramètres en ce qui concerne notre vie privée ? S'il n'y avait qu'eux, il n'y aurait pas vraiment lieu de s'inquiéter. Le problème naît de l'association de ceux-ci avec les autres informations que le serveur a pu glaner sur l'utilisateur et le lien qu'il peut éventuellement faire avec les cookies ou les fichiers d'audit (qui peuvent conserver une trace de la configuration de la machine de cet utilisateur d'une session à l'autre). Remarquons aussi qu'il est plus difficile, mais pas impossible, de se prémunir de l'envoi à notre insu de ce type d'informations (à savoir celles échangées par les protocoles TCP/IP et HTTP) que de l'utilisation d'autres techniques de collecte de données personnelles, comme nous le constaterons dans le deuxième chapitre à propos des navigateurs les plus courants.

[16] Il s'agit de la liste des cookies associés par le client HTTP à la ressource consultée.

I.3. Les Cookies

I.3.1. Petite présentation

Les cookies sont apparus à partir de 1995 avec la version 2 de Netscape qui les a inventés et déployés. Un cookie est un paquet de données (une chaîne de caractères), contenant des informations qui peuvent être compréhensibles ou non, transmis par un serveur visité (même indirectement, cfr. les hyperliens invisibles ci-dessous), à son initiative, au navigateur qui va le stocker de façon quasi permanente sur le disque dur de l'internaute (dans un fichier ou dans un répertoire particulier, au format texte banal), et généralement à l'insu de celui-ci. Le cookie étant identifié par le nom de domaine auquel appartient le serveur qui l'envoit, il sera dorénavant possible pour ce serveur de le modifier, le lire ou l'effacer à sa guise, et toujours sans que l'internaute s'en aperçoive. Sa durée de vie est variable selon la volonté du serveur qui en est à l'origine, cela va de quelques minutes (le temps d'une session) à quelques dizaines d'années. Le serveur en question peut le charger d'informations en fonction de sa connaissance des habitudes de navigation de l'internaute, changer sa date d'expiration ou même renvoyer des cookies supplémentaires puisque les navigateurs peuvent en contenir plusieurs centaines (jusqu'à 300 environ).

La propriété principale du cookie est qu'il est stable dans le temps (cette caractéristique est aussi son inconvénient principal vis-à-vis de la protection de la vie privée) et lié à une machine particulière (et donc à son utilisateur, s'il y en a qu'un seul qui emploie celle-ci). Ainsi, il permet à un site de se libérer du caractère dynamique que peut avoir l'adresse IP de ses visiteurs, ce qui lui permet donc de profiler une machine particulière pendant plusieurs années. En effet, à chaque fois que le surfeur reviendra sur le site, ce dernier saura qui est en face de lui et pourra réagir en conséquence (adaptation de l'information, publicités ciblées,...) avant même que la page Web désirée ne soit complètement téléchargée. C'est la raison pour laquelle les firmes de cyber-marketing vont utiliser systématiquement des cookies dans des hyperliens invisibles situés dans les bannières affichées sur leurs sites affiliés. C'est un des dangers majeurs d'atteinte aux données personnelles, car il donne à un site Web le moyen de garder trace des comportements et préférences d'un utilisateur.

Remarquons que cette technique n'a été rendue possible que par la collaboration des producteurs de logiciels de navigation qui ont programmé ceux-ci pour qu'ils puissent recevoir et gérer ces cookies.

I.3.2. Définition et utilisation

Un cookie (dont la taille ne dépasse pas les 4 Ko) est composé des informations suivantes[17] :

- son nom – cet attribut est obligatoire (un même serveur peut stocker jusqu'à 20 cookies maximum, ceux-ci ayant des noms différents)
- sa valeur (la plupart du temps inintelligibles par l'internaute)
- sa date d'expiration – cet attribut permet de définir la date à laquelle le cookie ne doit plus être stocké sur le disque, et ne doit plus être pris en compte par le serveur (souvent éloignée). Les cookies sans durée

[17] Les cookies sont définis par la RFC 2109 (première définition), http://www.w3.org/Protocols/rfc2109/rfc2109 (cfr. [22]), et RFC 2965, http://www.w3.org/Protocols/rfc2965/rfc2965 (cfr. [24]).

déterminée sont appelés des cookies de « session » et disparaissent lorsque l'on quitte le navigateur ou que la connexion réseau avec le serveur se ferme.

- le nom du domaine auquel appartient le serveur qui a envoyé le cookie - une machine provenant d'un domaine spécifique ne peut spécifier qu'un nom de sous-domaine ou son propre nom de domaine (seuls les ordinateurs du même domaine peuvent accéder à ce cookie, ce qui permet d'éviter qu'un cookie soit lu par un site concurrent)
- d'autres données techniques comme, par exemple, sa version ou les conditions de sécurité (envoi du cookie en protocole HTTPS (Secure HyperText Transfer Protocol), qui est une version sécurisée de HTTP s'appuyant sur le protocole SSL[18]).

Tant que sa date d'expiration n'est pas périmée, le navigateur communiquera le cookie systématiquement lors de chacune de ses requêtes (que ce soit suite à un clic sur un hyperlien visible ou suite à l'action d'un hyperlien invisible) si le serveur appartient au même domaine DNS que le serveur ayant transmis le cookie initial. C'est dans sa réponse à la requête que le serveur pourra modifier les informations contenues dans le cookie, le supprimer ou en envoyer d'autres. Les cookies font partie intégrante du trafic HTTP[19].

Les cookies peuvent être utilisés pour remplir des objectifs très variés. Voici quelques exemples typiques d'utilisations réalisables :

- Un serveur d'actualité ou d'articles de presse demande à ses utilisateurs de remplir un formulaire pour indiquer leurs préférences, afin de les stocker dans un cookie sur leur ordinateur, pour leur présenter directement, lors de leurs prochaines visites, les informations correspondant à ces préférences. Ainsi, ils ne devront répondre qu'une seule fois aux questions posées dans le formulaire, ce qui leur évitera d'avoir à y répondre chaque fois qu'ils reviennent sur le site. Le cookie peut aussi mémoriser la date de leur dernière visite sur le site afin de leur présenter uniquement la liste des dernières nouveautés.
- Un serveur de commerce électronique insère un cookie à chaque fois qu'un utilisateur sélectionne un produit, au même rythme que celui du remplissage d'un panier d'achat. Lorsque l'utilisateur se rendra sur la page contenant le formulaire de commande, le serveur récupérera l'ensemble des cookies du panier afin d'afficher l'ensemble des produits sélectionnés par le client.

---

[18] Le SSL (Secure Socket Layer), protocole de communication sécurisée standard utilisé par les navigateurs, permet à la fois le chiffrement lors de l'envoi de données sur Internet (typiquement un numéro de carte de crédit) et l'identification du site visité (par le biais d'un tiers de confiance) tout en garantissant l'intégrité des données échangées au cours de leur transport.

[19] Comme nous l'avons décrit supra, l'échange de cookies entre un client et un serveur se fait généralement de la façon suivante : un champ SET-COOKIE est placé dans l'en-tête d'une réponse HTTP, et le renvoi du cookie prend alors la forme d'un champ COOKIE dans le bavardage du navigateur. Mais techniquement, il est également possible de placer des cookies par l'intermédiaire de codes JavaScript (cfr. section suivante) ou à l'intérieur des champs <META-HTTP EQUIV> qui se trouvent dans le code HTML de la page Web elle-même. Ces champs, appelés meta-tags sont des balises HTML qui donnent des informations sur une page web. Contrairement aux tags HTML habituels, les meta-tags n'ont pas d'incidence sur la manière dont la page est affichée. En revanche, ils donnent des informations sur le créateur de la page, le nombre de mises à jour, son sujet et les mots clés caractérisant le contenu de la page. De nombreux moteurs de recherche emploient ces informations pour construire leur index.

- Un portail[20] ajoute des cookies en fonction des thèmes visités par l'internaute pour afficher dynamiquement des bandeaux publicitaires correspondant aux centres d'intérêts de l'utilisateur, et qui auront été découverts de cette manière. Les cookies mémorisent aussi le bandeau publicitaire actuellement affiché ce qui permet au serveur de publicités de changer efficacement de publicité à intervalles réguliers.
- Un serveur propose à l'utilisateur de choisir la configuration visuelle du site (la présence de cadres, d'animations ou encore la version linguistique souhaitée par exemple). Il enregistre toutes ces informations dans un cookie et présente à l'internaute une page d'accueil correspondant exactement à ses envies.
- Un serveur Web d'un grand événement sportif utilise les cookies pour aiguiller l'utilisateur vers le site miroir le plus proche de son lieu de connexion.
- Un site peut aussi se servir d'un cookie pour indiquer la date et l'heure de la dernière visite d'un internaute qui l'a consulté, ou un compteur qui mesure la fidélité de cet internaute à ce site.
- Un service de mail gratuit ou Webmail peut placer un cookie sur l'ordinateur d'un de ses utilisateurs afin de l'identifier à chacune de ses visites, ce qui fait qu'il ne doit plus s'identifier chaque fois qu'il vient voir s'il a reçu du courrier. Le site peut alors afficher une page d'accueil personnalisée avec son nom et l'utilisateur n'a plus qu'à saisir son mot de passe pour accéder à sa boîte de réception de messages.

Ces différentes utilisations des cookies peuvent évidemment être combinées entre elles à partir du moment où un site offre à ses visiteurs plusieurs services distincts en même temps (pratique courante sur de nombreux sites portails aujourd'hui). Mais ces différents cas de figure montrent surtout un des grands avantages des cookies : ils dispensent d'un stockage sur le serveur. C'est pourquoi le cookie est un outil technique parfois nécessaire à la navigation puisqu'ils permettent la conservation d'une information temporaire (ce qui n'est pas possible dans le cadre de l'utilisation du protocole HTTP[21]). Mais leur inconvénient principal est alors que, si l'internaute utilise un autre ordinateur, les cookies de sa machine usuelle n'y seront pas

---

[20] Un site portail fournit une vue d'ensemble de liens Web d'une manière ordonnée ou thématique. Les portails modernes sont des "super-sites" proposant toute une gamme de services, notamment la recherche sur le web, des nouvelles, les annuaires pages blanches et jaunes, des e-mails gratuits, des groupes de discussion, du shopping en ligne et des liens vers d'autres sites.

[21] Le problème est que le protocole HTTP est « stateless » (sans état), c'est-à-dire que chaque requête est indépendante des requêtes précédentes et futures, le serveur ne gardant pas la mémoire des états et étant donc incapable d'associer une série d'opérations à un navigateur particulier. Ce protocole n'offre donc pas de mécanisme pour être sûr que la page consultée est le résultat d'une navigation selon un certain ordre donné. De même, le HTTP est incapable de gérer la notion de session (une session étant l'intervalle de temps qui sépare une connexion d'une déconnexion). Le cookie est un artifice qui permet de pallier à l'effet « stateless » puisque le cheminement sur le site pourra être enregistré grâce à lui. D'une manière générale, tous les sites qui demandent une identification au démarrage (pour des raisons de sécurité et/ou d'authentification, par exemple) doivent gérer la notion de session, afin que l'utilisateur soit reconnu par le serveur de manière unique. Comme le HTTP ne la gère pas, il faut lui ajouter un petit « quelque chose », à savoir le cookie. Le principe est alors le suivant : les données de session (typiquement, le login et le mot de passe) sont stockées quelque part sur le serveur, et sont repérées par un identifiant unique. Cet identifiant est codé dans le cookie. Lequel cookie est renvoyé avec la page qui suit celle de l'identification et le navigateur le stocke sur le disque lorsqu'il le reçoit. Ensuite, à chaque demande de page sur le site, le navigateur enverra le contenu du cookie vers le serveur qui reconnaîtra l'utilisateur et pourra donc lui fournir les données qui lui sont associées. Un cookie est nécessaire pour identifier univoquement un utilisateur puisque l'adresse IP reste un identificateur relativement peu fiable, du fait qu'il peut être occulté par des serveurs intermédiaires (des serveurs proxies ou des firewalls), et que son caractère est souvent dynamique. Ce qui explique pourquoi les cookies (de session) sont, dans certains cas, indispensables à la navigation.

présents. De même, le serveur ne peut jamais être certain que le fichier qui contient le cookie n'aura pas été effacé ou modifié intentionnellement par l'internaute lui-même. C'est pourquoi les cookies ne devraient normalement servir que pour une unique utilisation au cours d'une seule et même session, mais c'est loin d'être la règle puisque les cookies de session sont moins répandus que les cookies ayant une longue durée de vie !

Les informations figurant dans le cookie peuvent être compréhensibles ou non. C'est le serveur qui l'envoie qui décide. Cela peut, par exemple, être un code qui renvoie à des informations stockées sur le site. Ainsi lorsqu'un internaute remplit un formulaire, le serveur peut enregistrer les informations communiquées sur son propre système de stockage sous un certain numéro de référence et stocker ce numéro dans un cookie qu'il communique au navigateur de cet internaute. Lors ces visites suivantes, ce serveur ne doit plus lui demander son identité tout en faisant le lien avec lui. Par exemple, ce numéro pourrait être un index dans une base de données située chez une firme cyber-marketing et qui contiendrait les informations personnelles que cette firme possède à propos de l'internaute. Si les informations contenues dans le cookie sont chiffrées, c'est aussi pour les rendre difficiles à manipuler, par l'internaute lui-même par exemple, ou pour les rendre difficiles à exploiter par un tiers extérieur, ou encore pour les compresser afin qu'elles prennent moins de place. Hors, les données contenues dans un cookie peuvent aussi être « sensibles » (nom d'utilisateur, mot de passe ou d'autres informations fournies lors de l'inscription à un site, afin d'éviter d'avoir à les retaper à chaque fois), données personnelles (numéros de cartes de crédit, ...), données confidentielles (statut médical, tendance politique, identité culturelle ou sociale...) ou tout autre donnée que l'internaute aurait communiqué à un site de son plein gré.

I.3.3. Leurs menaces vis-à-vis de la vie privée

Jusqu'à présent, les cookies représentent la seule manière dont un site peut mémoriser des informations associées avec un internaute particulier, les stocker sur son disque dur et les relire à chaque interaction avec le site. Le site ne pouvant placer les cookies que par l'intermédiaire des navigateurs, ceux-ci disposent fréquemment d'une option permettant de demander l'autorisation de son utilisateur avant d'enregistrer un cookie. Mais une fois que le site a placé un cookie, il pourra le lire à l'avenir sans sa permission. Bien sûr, un site ne peut lire que les cookies qu'il a stockés et non ceux enregistrés par d'autres sites et vice-versa.
Heureusement, un site ne peut pas écrire d'informations en un endroit arbitraire de notre disque dur, les cookies étant stockés dans un fichier spécifique propre à chaque navigateur. En outre, le site ne peut rien y écrire sans notre permission à condition que le browser soit configuré pour agir ainsi. De même, le site ne peut pas lire n'importe quelles informations de notre disque.

Comme nous l'avons décrit, si un site peut stocker un cookie, il peut garder une trace de tout ce qu'un internaute a fait en enregistrant simplement ces informations dans un cookie qu'il continue à mettre à jour continuellement. Il peut ainsi élaborer son profil : cela peut être positif ou négatif selon l'utilisation de ces informations par le site. Or lorsqu'un site place un cookie, il est le seul à pouvoir ensuite le lire. Par conséquent, cela lui permet d'établir un profil de son comportement pendant que l'internaute le visite mais pas de son comportement général lorsqu'il surfe sur le Web. Il a donc au moins l'impression que les données rassemblées sur lui sont spécifiques au site et que personne ne peut établir un profil complet le concernant.
Supposons maintenant que, pendant que nous visitons le site *www.actualités.com*, un cookie soit placé, non pas par ce site en question mais par un autre diffusant des publicités,

appelons-le *www.marketing.com*, qui est le site d'une société de cybermarketing. Le site *www.actualités.com* peut effectuer cela très simplement en affichant une image provenant de *www.marketing.com* sur sa page d'accueil. Ainsi, pendant que nous visitons *www.actualités.com*, notre navigateur émet sans nous avertir une requête vers le site *www.marketing.com* pour récupérer l'image, et celui-ci en profite pour placer un cookie sur notre disque dur. Supposons que *www.marketing.com* ait enrôlé de nombreux autres sites pour afficher ses images et transmettre des cookies à leurs visiteurs. *www.marketing.com* établira ainsi, grâce aux cookies envoyés, un profil contenant l'ensemble des informations récoltées sur notre comportement général lors de notre visite sur ces différents sites. Plus le nombre de sites enrôlés par *www.marketing.com* pour afficher ses images est élevé, plus le profil nous concernant sera complet et précis. Les cookies envoyés par un site autre que celui que nous pensons visiter sont appelés cookies étrangers ou cookies tiers.

De plus, grâce aux cookies, les serveurs Web ont la capacité d'attribuer à chaque ordinateur qui rentre en contact avec eux un numéro particulier servant d'identifiant unique global (numéro souvent indépendant de l'adresse IP de cet ordinateur), c'est ce qu'on appelle un GUI (Global Unique Identifier). Pour parfaire les analyses faites par les outils de profilage, un cookie enregistrera donc cet identifiant qui permettra de faire le rapprochement entre les différentes adresses IP attribuées à cet internaute par son fournisseur d'accès au cours de ces différentes connexions à Internet[22]. C'est ainsi que la précision des profils établis par ces outils a pu progresser qualitativement. En effet, toutes les données récoltées à propos d'un internaute peuvent être amalgamées pour créer un profil général de celui-ci, et cela grâce à l'identification unique contenue dans le cookie. Il ne reste plus qu'un risque d'erreur résiduel : dans le cas où, pour un même ordinateur, plusieurs personnes physiques différentes se connectent habituellement.

Mais, les cookies peuvent aussi permettre à un serveur de déterminer le parcours d'un internaute durant sa visite mais aussi la fréquence de ses visites, le nombre de pages consultées,..., ce qu'on appelle son « clickstream », et d'établir son comportement, conscient ou inconscient, sans avoir besoin d'utiliser des fichiers d'audit ou une autre technique plus exigeante à mettre en oeuvre. Pour ce faire, le serveur doit juste introduire un cookie à chaque page ou lors de chaque action particulière que l'internaute exécute. Ensuite, il les récupère tous ensemble pour obtenir aisément l'entièreté du chemin parcouru ou des actions exécutées[23]. Enfin, rien ne l'empêche de proposer à l'internaute en question, lors de ses prochaines visites, des pages créées dynamiquement en fonction de ce profil qu'il a établi.

Souvent, ceux qui sont à l'origine des cookies et qui gèrent leur contenu sont non pas les auteurs du site mais plutôt ses partenaires commerciaux, les régies publicitaires et les sociétés de sondage essentiellement, et ce toujours pour la même raison, connaître les centres d'intérêts propres à chaque internaute en particulier. Cette information est acheminée vers les serveurs des grands sites gérant l'affichage des annonces publicitaires, et cet internaute recevra alors des publicités ciblées lorsqu'il consultera ses sites préférés. D'ailleurs, d'énormes progrès en matière de « profiling » ont été accomplis afin de réaliser des références croisées entre tous les sites et toutes les pages qu'un internaute a visités afin d'améliorer

[22] N'oublions pas que, dans de nombreux cas encore, l'internaute se connectant à Internet via son (ou même ses) fournisseur(s) d'accès reçoit une adresse IP différente à chaque fois.

[23] C'est pourquoi les cookies sont considérés par de nombreuses personnes comme des « mouchards ».

encore et toujours la qualité de l'impact des publicités visionnées par cet individu, triste consommateur qu'il est poussé à devenir (ou à rester) !

Quant aux différents sites que cet internaute a visités, qu'ils soient contrôlés par des amateurs ou des professionnels, ils se font rémunérer aux "clics" que l'ensemble de leurs visiteurs effectue sur les bandeaux publicitaires[24]. Cette pratique est réalisée des millions de fois par jour par des sites vendant leur espace Web pour afficher les bannières de pub des firmes de cyber-marketing. Il faut peut-être rappeler que la gratuité sur Internet n'existe quasiment pas, il s'agit juste d'un nouveau modèle économique. Si aujourd'hui nous pouvons visiter des millions de sites et lire des milliards de pages gratuitement et en toute liberté, c'est parce que les sondages et la publicité rémunèrent les « créateurs de contenu ». En effet, le marché des données est en plein essor sur Internet, et la valeur boursière des opérateurs de sites est directement liée au nombre des visites et aux traces qui sont laissées. L'information accumulée sur les visiteurs a donc de la valeur financière. Ceci explique pourquoi l'utilisation du cookie est un enjeu commercial, puisqu'il permet aux sites commerciaux et aux régies publicitaires du Web de compiler en temps réel des informations que des centaines d'études marketing ne pourraient leur fournir que mois après mois. Grâce aux informations connues et stockées dans les cookies, on établit des profils des internautes : profil commercial, politique, religieux, culturel, social, médical. Leur utilisation peut être commerciale ou parfois même un instrument de censure gouvernementale ou privée (dans le cadre d'un salarié, par exemple).

I.3.4. Faut-il s'en méfier ?

Mais comment peut-on distinguer le site qui utilise les cookies pour améliorer le « confort » de notre consultation de ceux qui les emploient pour nous traquer et nous pister à notre insu ? En réalité, un cookie n'a rien de dangereux en soi. Les informations contenues dans un cookie ne sont pas nominatives : elles identifient tout au plus le navigateur de l'internaute ou l'adresse IP de son ordinateur. Ainsi, un serveur ne peut pas disposer, par leur intermédiaire, d'informations qu'un internaute ne lui aurait pas précédemment communiquées de son plein gré. On en déduit donc que les cookies ne pourraient contenir son identité civile ou son adresse e-mail, que s'il les avait donné explicitement[25] (dans un formulaire par exemple). C'est uniquement à partir de ce moment qu'il sera possible de lier un identifiant déposé dans le cookie avec les informations personnelles de l'internaute. En outre, seul l'ensemble des serveurs appartenant au même domaine que celui qui a donné cette information pourrait l'exploiter plus tard, à moins bien sûr que cette information ne soit échangée ou revendue à des tiers. En fait, un cookie contient le plus souvent : la date et l'heure d'une visite, et/ou un compteur qui mesure la fidélité à un site.

Le problème des cookies est lié à leur non transparence : leur contenu étant rarement explicite, on ne peut pas savoir à quoi ils peuvent servir (sont-ils là dans un but de profilage ou pour améliorer le confort de notre visite ?). Au-delà de l'impression de « manipulation » que cette technique peut produire, et même en sachant qu'un cookie ne peut pas contenir

[24] D'ailleurs, les sites utilisant cette pratique ignorent parfois que cette dernière est capable de communiquer autant d'informations sur leurs visiteurs.

[25] Remarquons que c'est le système de « l'opt-out » qui est utilisé par défaut sur de nombreux sites, à savoir que l'accord implicite de l'internaute suffit à l'exploitation des données qu'il a communiquées de son plein gré. Il est regrettable que ce système soit la plupart du temps d'application au lieu de « l'opt-in », à savoir la présomption strictement inverse.

grand-chose en lui-même ni avoir d'action particulière, on reste sceptique à l'idée qu'un serveur distant puisse placer des informations sur l'ordinateur d'un internaute sans avoir particulièrement à s'expliquer sur son contenu ni sur l'utilisation qu'il en fera.

Quoique nous fassions sur un site particulier, n'oublions pas que tous les cookies enregistrés auparavant et auxquels nous n'avons pas prêté attention pourront être mis en relation avec l'information beaucoup plus personnelle que nous pourrions lui transmettre à n'importe quel moment. Il vaut la peine de souligner que la collecte de données sur les internautes n'entraîne généralement aucun coût pour l'entreprise, du fait que les consommateurs fournissent souvent ces informations eux-mêmes, par exemple en remplissant un questionnaire. Les sites Web utilisent souvent des programmes de fidélisation tels que des jeux, questionnaires, nouvelles, qui impliquent que leurs visiteurs fournissent des informations personnelles.
C'est pourquoi il est important que les internautes apprennent à maîtriser et à contrôler (voire éliminer) leurs traces sur la toile[26] (en ne divulguant qu'un minimum d'informations, en étant vigilant quant aux actions qu'ils exécutent, en s'informant régulièrement,...). Les sociétés de cybermarketing raffolent des informations personnelles suffisamment précises car elles permettent d'établir plus rapidement le profil d'une personne : il s'agit généralement d'une réponse à un questionnaire ou à un formulaire. Dans la plupart des cas, c'est une information volontairement et innocemment transmise au site visité, qui multiplie les sondages et autres quiz pour, par exemple, faire gagner des bons de réduction. Le plus souvent, l'objectif de ces questionnaires en ligne pour le site qui les propose est de mieux cerner les attentes des internautes vis-à-vis de leur contenu, mais parfois, ils serviront plutôt à établir un profil plus précis des habitudes de consommation de leurs visiteurs.

Faut-il avoir peur des cookies ? Non, le cookie est un outil technique (permettant d'utiliser de façon pratique la notion de session, par exemple) parmi tant d'autres. Il est régi par les mêmes règles au sein de tous les navigateurs qui permettent leur utilisation. En lui-même le cookie n'est pas dangereux, c'est son utilisation qui peut l'être. On doit admettre que beaucoup de choses exagérées ont été écrites sur les menaces que font peser les cookies sur la vie privée des utilisateurs : reconnaissons que dans l'ensemble, ils ne contiennent pas de données particulièrement identifiantes ni sensibles. Toutefois, on ne saurait écarter l'hypothèse que des traitements et des échanges de données opérés par des sites permettent au bout du compte de mettre en relation des cookies avec des bases de données de profils-utilisateurs beaucoup plus riches. Rien n'interdit non plus d'introduire dans un cookie l'adresse e-mail qu'un utilisateur a bien voulu confier à un site en remplissant un formulaire. On ne peut exclure non plus, que des sites distincts mais non moins alliés s'accordent techniquement de manière à ce que leurs cookies soient inter-opérables. Il importe donc de savoir comment s'en préserver lorsqu'ils ne semblent pas être utiles, d'être vigilant quant aux informations que nous transmettons sur Internet, et d'exiger des sites qui exploitent les cookies plus de transparence quant à l'utilisation qu'ils en font réellement.

Il convient également de mentionner que, même si l'on a désactivé les cookies, le site peut toujours assurer notre suivi, du moins pendant que nous le visitons. En voici un exemple : le site peut stocker les informations non dans un cookie sur notre machine mais dans les liens qui nous sont accessibles. Chaque lien disponible contient l'adresse de la prochaine page à récupérer. Toutefois, le site peut personnaliser spécifiquement le lien de façon à ce qu'il contienne également des informations de suivi.

[26] Nous invitons le lecteur à parcourir le deuxième chapitre pour davantage d'informations à ce sujet.

Pour plus de clarté, supposons que nous visitons un site appelé *www.suivi.com*, qui présente une page d'accueil contenant un lien vers une autre page. Nous voyons à l'écran du texte décrivant le lien (par exemple, « visitez notre deuxième page »). Outre le texte visible, le lien contient également l'adresse de la deuxième page, par exemple *www.suivi.com/secondepage*. Supposons que le lien de la page d'accueil ne contienne pas uniquement *www.suivi.com/secondepage* mais plutôt quelque chose comme *www.suivi.com/secondepage?0*. Le texte « *?0* » peut être un code indiquant que nous n'avons pas encore visité la deuxième page. Supposons que nous cliquons sur ce lien et affichons la deuxième page. Ensuite, nous cliquons sur un lien de la deuxième page nous ramenant à la page d'accueil. Cette dernière est différente de la page d'accueil affichée la première fois parce que le lien vers l'adresse *www.suivi.com/secondepage* contient désormais le code *www.suivi.com/secondepage?1*. Le site utilise alors la page proprement dite (et non un cookie) pour assurer le suivi des pages visitées et des éléments sélectionnés.
La bonne nouvelle est que ce type de suivi ne fonctionne que tant que nous naviguons dans le site et visitons ses pages connexes. Toutes ces informations sont perdues lorsque nous quittons le site. Si nous y revenons ultérieurement, le lien *www.suivi.com/secondepage?0* sera de nouveau affiché. Evidemment, si nous plaçons une page d'un tel site dans nos signets et que nous y retournons par ce biais, les informations de suivi seront toujours présentes.

I.4. Le scripting côté client : JavaScript, VBScript, ActiveX et applet Java

Le JavaScript est un langage de scripts qui s'insère dans le code HTML d'une page Web. Il a été mis au point par la compagnie Netscape en 1995[27] et peut être utilisé (reconnu et interprété) avec la majorité des navigateurs actuels. L'utilisation de JavaScript ajoute de l'esthétisme et de la convivialité. Ainsi, il permet aux auteurs de pages Web de créer des animations et de rendre la consultation plus interactive pour les usagers avec par exemple une horloge, un texte qui défile, une image qui change au passage de la souris, des formulaires et de nombreuses autres fonctionnalités pour améliorer plus encore le contenu des pages. Ces améliorations au langage HTML permettent d'exécuter des commandes du côté client, c'est-à-dire au niveau du navigateur et non du serveur web. Ainsi le JavaScript est dépendant du navigateur appelant la page Web dans laquelle le script est incorporé.

Bien que les noms soient similaires, il ne faut pas confondre le langage Java avec le JavaScript. Java est un langage de programmation utilisant l'architecture client/serveur. Les applications Java sont préalablement compilées par une machine virtuelle. Ce langage orienté objet est un langage de haut niveau, complet, stable et multi-plateforme. Mais il est aussi très utile pour créer des mini-applications dynamiques (les applets[28]) pour l'Internet, les intranets

[27] A l'origine, il se nommait LiveScript et était destiné au navigateur Netscape Navigator 2. Historiquement il s'agit même du premier langage de script conçu spécialement pour le Web. Suite à une association avec le constructeur Sun, Netscape rebaptise son langage Javascript (un clin d'oeil au langage Java développé par Sun dont il est proche du point de vue de la syntaxe). A la même époque, Microsoft met au point le langage Jscript, un langage de script très similaire mais qui n'est pas 100% compatible avec JavaScript. Ainsi, pour éviter des dérives de part et d'autre, un standard a été défini pour normaliser les langages de script, il s'agit de l'ECMA 262, créé par l'organisation du même nom (ECMA, European Computer Manufactures Association). Depuis, le JavaScript a été très largement utilisé du fait de ses fonctionnalités avantageuses qui sont au-delà du langage de description de pages HTML.

[28] Une applet (APPlication Light wEighT) est une petite application indépendante de la plate-forme (compatible Windows ou Macintosh, etc) et du navigateur utilisés, qui est téléchargée depuis un serveur du WWW et qui est exécutée localement au sein d'un navigateur.

et autres réseaux distribués. La plupart des navigateurs récents possèdent un interpréteur Java intégré. Celui-ci est en fait une machine virtuelle (un programme adapté au système d'exploitation de la machine cliente et qui fait le lien entre les fonctions appelées par l'applet et les fonctions du système d'exploitation) qui permet d'interpréter le pseudo-code (l'applet compilée) et qui doit être chargée en mémoire (du côté du client) à chaque chargement de la page, d'où un important ralentissement pour les applets Java contrairement au JavaScript.

Le JavaScript[29], lui, est un langage interprété basé objet dont le code est directement écrit dans la page HTML, c'est un langage relativement peu évolué qui ne permet aucune confidentialité au niveau des codes (ceux-ci sont clairement visibles). Remarquons qu'il permet également de créer, modifier et détruire les cookies déposés dans un ordinateur. Il existe aussi un autre langage de script, nommé VBScript[30], qui offre en gros les mêmes possibilités que le JavaScript. Bien qu'il soit plus pauvre, moins autonome et beaucoup moins répandu que son concurrent, il prend tout son intérêt lorsqu'il est utilisé avec la technologie ActiveX.

Java	JavaScript
Langage pseudo-compilé, le code source étant compilé avant son exécution (chargement d'une machine virtuelle)	Langage interprété par le navigateur au moment de l'exécution
Code (applet) à part du document HTML appelé à partir de la page	Code intégré au HTML
Pas accès aux objets du navigateur	Accès aux objets du navigateur
Langage de programmation de haut niveau	Code de programmation simple mais limité
Confidentialité du code	Accessibilité du code

ActiveX de Microsoft est similaire à JavaScript car il est aussi interprété par le navigateur. Il est basé sur le Component Object Model de Microsoft, qui permet de visualiser des fichiers des autres applications Microsoft, tout en permettant aussi d'ajouter de l'esthétisme et de l'interactivité aux pages Web[31].

---

[29] Coté client, il faut avoir un navigateur Netscape (vers. 2 pour JavaScript 1.0, vers. 3 pour JavaScript 1.1, vers. 4.0 à 4.05 pour JavaScript 1.2, vers. 4.06 à 4.7 pour JavaScript 1.3...) ou MS Internet Explorer (à partir de vers. 3 pour JavaScript 1.0, vers. 4 pour JavaScript 1.2, vers. 5-5.5 pour JavaScript 1.3…). Pour obtenir des codes parfaitement compatibles il est donc nécessaire de connaître précisément la version du navigateur qu'un visiteur utilise lors de sa consultation d'un site, c'est pourquoi des codes JavaScript permettent de « renifler » les variables d'environnements des navigateurs.

[30] En fait, il a été créé par Microsoft afin de concurrencer le JavaScript de Sun et Netscape tout en étant dérivé de leur langage propriétaire, le Visual Basic, dont il hérite de la syntaxe. Coté client, il faut avoir un browser MS Internet Explorer (vers. 3 pour VBScript 1.0, vers. 4 pour VBScript 2.0...) ou un browser Netscape doté d'un plug-in adéquat.

[31] Tout a commencé au début des années 90 avec l'apparition de DDE (Dynamic Data Exchange) dans les premières versions de Windows. Il s'agissait à l'époque de permettre l'échange de données entre deux applications Windows. Quelques temps plus tard, durant l'année 91, Microsoft propose la 1ère version d'un mécanisme plus évolué que DDE, appelé OLE (Object Linking and Embedding) qui étend les capacités de communication inter-applications en introduisant une approche objet. OLE permettait d'intégrer, dans une

Les deux composants technologiques de base présents au sein des browsers, qu'il s'agisse de scripts, de Java ou de ActiveX, s'avèrent presque toujours porteurs d'anomalies qui menacent la sécurité des données et des fichiers personnels présents sur les ordinateurs des utilisateurs.
Ainsi, il existe un trou de sécurité dans l'implémentation du JavaScript sur la version 2 des navigateurs Netscape, qui a été rectifié dans les versions suivantes de ce navigateur, qui permettait à un site Web malintentionné d'être en capacité de capturer l'adresse e-mail d'un internaute à son insu (un formulaire avec du code JavaScript malicieux permettrait l'envoi dissimulé d'une adresse e-mail sans que l'internaute ne la délivre explicitement).

Mais quels sont les dangers liés au code malveillant ? Il est impossible d'évaluer même approximativement l'étendue des attaques délibérées et malveillantes perpétrées sur Internet. On relève cependant de nombreux cas d'utilisation de code favorisant les attaques malveillantes ou intentionnellement conçu pour les déclencher.
En voici trois manifestations principales :

- ActiveX : les applications développées en ActiveX peuvent disposer d'un accès intégral au disque dur du système de l'utilisateur. Nul n'est donc besoin d'énumérer les risques liés à une utilisation malintentionnée de cette technologie. Particulièrement préoccupants sont les cas de code ActiveX malveillant capable de dérober des données à l'insu de leur propriétaire ou de détruire des fichiers sur les ordinateurs des utilisateurs, voire de lancer une commande de formatage du disque dur[32].
- Java : menace potentielle similaire à celle du code ActiveX, les applets Java permettent d'effectuer des accès illicites sur le système de l'utilisateur[33].
- JavaScript et VBScript : ils sont capables de déposer des informations chez l'utilisateur ou de prélever des données pour les transférer sur le serveur hôte, à l'insu de cet utilisateur. Les scripts sont activés lorsqu'une action bien précise

---

application, des objets issus d'autres applications Windows. Cependant, OLE 1.0 souffrait de certaines limitations et n'était pas très performant. C'est pourquoi Microsoft sortit deux ans plus tard la 2ème version d'OLE, OLE 2.0, corrigeant non seulement les défauts de la première mouture d'OLE mais introduisant également le modèle COM (Component Object Model) qui définissait le format binaire interne des objets OLE. En 1994, avec le passage de Windows au 32 bits, COM devint la base même de toute l'architecture OLE. Microsoft introduit alors la notion de composant OLE ou OCX (OLE Control eXtension), basé, entre autre, sur les extensions VBX de VisualBasic. C'est sur ces composants qu'est bâtie la technologie ActiveX introduite début 96 par Microsoft et qui constitue le cœur de sa stratégie Internet.
Dans la terminologie ActiveX, on ne parle plus d'application mais de document. Un document est un réceptacle à contrôles ActiveX. Un contrôle ActiveX est un élément logiciel distinct qui respecte le modèle COM. Un contrôle peut communiquer avec le document qui le contient, et ce, de façon bidirectionnelle. ActiveX a été spécialement conçu pour fonctionner dans un environnement World Wide Web, grâce notamment à Microsoft Internet Explorer (MSIE) qui est un des rares navigateurs à pouvoir recevoir des contrôles ActiveX.
Ces contrôles sont intégrés dans une page web grâce à une balise HTML dédiée. Un langage de type script, le VBScript justement, permet d'agir sur les contrôles ActiveX intégrés à une page web. Ainsi il peut servir de « colle » entre le monde Windows et le Web.

[32] Le groupe de hackers allemands *Chaos Computer Club* a réussi à démontrer qu'un contrôle ActiveX pouvait déclencher une transaction non-autorisée sur le compte bancaire d'un utilisateur à partir du logiciel de comptabilité personnelle Quicken.

[33] Le *Brown Orifice Vulnerability* est un trou de sécurité présent dans les anciennes versions de Netscape Communicator (de 4.4 à 4.74) et qui permet à quiconque sur Internet d'utiliser ces navigateurs pour accéder à la totalité des fichiers contenus sur le disque de l'utilisateur.

d'un utilisateur est effectuée (le plus souvent un clic sur un lien hyperlien ou un bouton, ou simplement au chargement d'une page). Ces scripts sont exécutés en temps réel et pourraient avoir certains accès douteux (la sécurité des navigateurs ayant dû être revue plusieurs fois[34]). Le plus ennuyeux, mais sans conséquence, est le phénomène des 'pop-ups', à savoir des pages (généralement ayant un format plus petit, mais pas forcément) qui s'ouvrent automatiquement et intempestivement lors du chargement d'une pages Web. De plus, ces langages de scripts permettent de se libérer du filtrage que peut effectuer un serveur proxy[35] au niveau des cookies et du masquage de l'adresse IP (pour en savoir plus sur les différents usages d'un serveur proxy, cfr. Chapitre suivant).

Sur Internet, les trous de sécurité sont fréquents, mais en plus toute personne qui le souhaite peut y accéder en lisant la large documentation disponible. Dans les cas les plus favorables, ces trous sont corrigés par un ou plusieurs correctifs (ce qu'on appelle souvent un « patch ») ou dans les versions ultérieures d'un même navigateur. Mais ils peuvent aussi être mal colmatés, faire l'objet de patchs défectueux, rester inconnus ou pis connus uniquement de certains groupes de hackers malveillants. Bien sûr, la majorité des internautes n'est pas avertie ou ignore tout simplement les dangers que ses données personnelles encourent lorsqu'il surfe avec son navigateur « troué ». C'est pourquoi, lors de chaque requête HTTP, l'envoi systématique de la signature du navigateur utilisé constitue une information utile pour le site Web malveillant qui pourra alors « choisir » un trou de sécurité adapté pour perpétrer ses méfaits. D'où le dilemme des alertes de sécurité : en révélant une faille, on la met à la disposition des hackers...

Comme nous l'avons évoqué antérieurement, une page Web peut contenir des programmes de scripting s'exécutant sur notre machine, ces programmes pouvant demander des informations sur notre machine et les renvoyer au site. Néanmoins, nous pouvons aisément faire en sorte de ne pas transmettre d'informations involontairement puisque l'autorisation par le navigateur d'exécuter le scripting côté client est contrôlée par certains paramètres de configuration. Par défaut, ces paramètres sont souvent configurés pour permettre l'exécution de ces scripts. En les modifiant, nous sommes capables d'empêcher le site de demander et de transmettre des informations (comprenant notamment le nombre (mais pas le nom) des sites précédemment visités, si notre navigateur peut exécuter des programmes écrits en langage Java, le nombre et le type de plugins installés dans le navigateur, les dimensions de sa fenêtre, lire et écrire des cookies, etc.).
Le scripting côté client est « normalement » incapable d'obtenir des informations personnelles réellement compromettantes. Toutefois, si le navigateur est mal configuré, le scripting peut collecter des informations beaucoup plus personnelles. En fait, il peut même lire des fichiers quelconques sur le disque dur et les renvoyer au site. La configuration par défaut de la plupart des navigateurs permet d'empêcher de telles actions de se produire en demandant l'autorisation explicite de l'utilisateur ou en l'interdisant (mais il existe des trous de sécurité permettant de contourner ces mesures de protection...). Avant de donner notre autorisation,

[34] Le « trou de Cuartango » en est un exemple célèbre.

[35] Ce type de serveur sert d'intermédiaire entre le surfeur et le site Web consulté. Il sert essentiellement d'antémémoire (cfr. Section sur la mémoire cache). Difficile de mettre en oeuvre des cookies dans ce cas de figure pour le site visité, car certains proxies sont capables de les filtrer, voire de les « manger ». Malheureusement, l'apparition des scripts a permis de contourner ce type de protection, ces derniers passant à travers comme si de rien n'était. En effet, les scripts sont incorporés à l'intérieur du code HTML des pages Web. Or le proxy ne filtre que les entêtes des réponses HTTP par leur contenu.

nous sommes averti que le site tente d'utiliser un code potentiellement dangereux. Il ne faut évidemment la donner que si l'on a une confiance absolue en ce site. Mais la meilleure façon de se protéger reste de désactiver toute utilisation de scripting au cours de la navigation et de veiller à maintenir à jour son browser au moyen des patchs de sécurité qui sont publiés régulièrement par les éditeurs (voire à changer de navigateur pour un autre réputé plus « sûr »). Malheureusement, supprimer l'action des scripts bien que tout à fait possible, via les options des navigateurs, a un effet désastreux sur la présentation des sites Web les utilisant : pages incomplètes ou pis vides de tout contenu.

I.5. La redirection automatique

Lorsqu'un serveur Web reçoit une requête HTTP portant sur une page Web, il peut répondre en indiquant un code de redirection vers une autre page Web (pouvant être une autre page du même site ou pouvant se trouver sur un autre serveur quelconque). Le navigateur recevant cette réponse va émettre une seconde requête afin d'obtenir la nouvelle page, en l'occurrence celle indiquée avec le code de redirection, et ce à l'insu de l'utilisateur, qui ne s'apercevra de rien. Cette méthode est utilisée par les firmes de cyber-marketing pour rediriger les clicks sur une de leurs bannières publicitaires (qui conduit à leur propre site) vers le site de leur client correspondant. C'est ainsi que l'utilisateur ne remarque pas qu'en ayant cliqué sur une bannière, il est d'abord passer par le site de la firme de cybermarketing qui a émis cette bannière avant d'atteindre le site de l'annonceur, à savoir la société dont la publicité vante les produits et/ou les services (c'est ce qu'on appelle le « *click-through* »). Il va de soi qu'au passage la firme de cyber-marketing a pu « (re)connaître » le surfeur ayant réagi au stimulus visuel de la pub pour mieux affiner son profil de consommateur.

I.6. Les hyperliens invisibles

Aujourd'hui, le concept d'hyperlien est connu de toutes les personnes ayant approché Internet de près ou de loin. On peut dire qu'il s'agit du concept fondamental au cœur de la navigation sur le réseau mondial. Ce qui n'est pas apparent à l'œil humain c'est que les navigateurs permettent d'inclure dans le code HTML d'une page Web un hyperlien permettant d'importer une image (cet hyperlien sera ouvert automatiquement par le navigateur). Cette image ne doit pas forcément être localisée sur le même serveur que celui qui a reçu la requête initiale, puisqu'elle peut très bien provenir d'un site tiers. Ainsi, contrairement aux hyperliens « classiques » (qu'on appelle aussi hyperliens visibles et explicites) qui requièrent une action voulue de la part de l'internaute, les hyperliens invisibles (et implicites) s'exécutent sans l'intervention de l'utilisateur et sans qu'il en soit au courant. C'est grâce à ce type d'hyperliens que les Webmasters peuvent inclure dans leurs pages des images provenant de sites tiers. Grâce à cette possibilité technique, les firmes de cyber-marketing peuvent diffuser leurs bannières publicitaires, stockées sur leur serveur, sur les milliers de pages appartenant aux différents sites supports se trouvant dispersés dans le monde. Dans le cas présent, la variable HTTP_REFERER contient la référence de la page transférante, c'est-à-dire la page principale dans laquelle les images seront localisées. En d'autres mots : si un site Web contient dans sa page en HTML un lien invisible vers une image située sur le site Web d'une

société de marketing électronique, cette société connaîtra la page transférante avant même d'envoyer sa bannière publicitaire[36].

Pour créer un hyperlien invisible, il suffit d'utiliser la balise suivante dans le code HTML d'une page Web, *<IMG src="*URL*">*, pour que le navigateur qui télécharge la page en question ouvre, de lui-même et sans en avertir son utilisateur, une deuxième session HTTP vers le site tiers spécifié dans l'URL pour charger l'image indiquée. De cette façon, l'utilisateur a l'impression que le contenu de la page s'affichant à l'écran est uniquement issu du site qu'il consulte. On peut donc affirmer que les hyperliens invisibles constituent le procédé par lequel le dialogue des navigateurs est envoyé à des serveurs tiers à la communication entre un internaute et le site qu'il visite. Pris isolément, ce procédé ne nuit que très peu à la protection de la vie privée de l'internaute.

Cas particulier : Les Web-Bugs ou Pixels Espions

Un pixel espion, en anglais Web Beacon ou Web Bug[37] (cfr. [28, 31]), est une image imperceptible incorporée abusivement dans une page Web, un message de courrier électronique (reçu avec n'importe quel logiciel de messagerie électronique, pourvu qu'il accepte les messages incorporant du code HTML et des JavaScripts) ou même un document réalisé avec une application bureautique ; elle est conçue pour identifier celui qui lit la page Web ou le courrier électronique (une newsletter ou un mailing publicitaire, par exemple). Le pixel espion est invisible et pratiquement indétectable, car sa taille n'est que d'un pixel sur un pixel (en fait, c'est un point coloré, mais ce point peut être rendu « transparent » s'il est de la même couleur que le fond de la page). Pourtant, il peut être relié à quelques lignes de code JavaScript qui lui confèrent des fonctions informatives. Ainsi dissimulé, il récupère des informations à l'insu de l'internaute telles que son adresse IP, le site qu'il a précédemment visité, son type de navigateur et/ou le temps qu'il a passé sur la page ainsi que les différents cookies précédemment mémorisés par son ordinateur, le tout pouvant finalement être stocké sur son propre disque dur sous la forme d'un autre cookie[38]. Cette capacité d'utilisation en ligne en fait l'outil idéal pour délivrer un cookie dans un contexte tiers. C'est pourquoi ce sont principalement des sites Web tiers, ceux de sociétés de cybermarketing, qui utilisent les données ainsi récoltées, plutôt que le site Web d'apparence anodine qui abrite les Web Bugs dans ses pages.
Ils sont également utilisés à des fins peu avouables par certains sites Web pour tenter de recueillir des informations à caractère privé, qu'ils transmettront, pour exploitation ultérieure, à un serveur distant.

Les utilisations des Web bugs sont nombreuses[39]. Lorsqu'ils sont placés sur une multitude de sites, il devient possible de suivre les déplacements d'un utilisateur en temps réel.

---

[36] Remarque : lorsqu'une recherche est faite par un moteur de recherche, le nom de la page Web contient les mots clés introduits.

[37] Le Web Bug est aussi appelé "clear GIF", "invisible GIF", "1-pixel GIF" ou encore "tracker GIF".

[38] Lorsque l'usage des Web Bugs est combiné avec celui des cookies, les possibilités sont plus étendues et on obtient un outil plus dangereux pour le respect de la vie privée, puisqu'une telle méthode permet au Web Bug de connaître, par l'intermédiaire des cookies, toutes les pages et tous les sites affiliés qui ont fait l'objet d'une visite à un moment donné par l'internaute et par là, son comportement en ligne précédent.

[39] Pour s'en faire une idée, http://www.privacyfoundation.org/resources/whyusewb.asp

Leur vocation est plus généralement de compter univoquement chaque visite sur une page, ce comptage étant effectué par des certificateurs d'audience[40]. Placé secrètement sur un site, ils permettent même d'en espionner l'audience. Selon la Privacy Foundation, le procédé serait déjà utilisé, en combinaison avec les cookies, par des sociétés de marketing et quelques certificateurs d'audience, pour mesurer, et suivre, précisément l'audience des publicités ciblées (c'est un compteur d'impact publicitaire et un outil très efficace pour établir des statistiques détaillées) et déterminer le profil des internautes (nos habitudes sur un site peuvent être étudiées et nos visites comptabilisées ainsi que nos heures de passage) à leur insu. Ils peuvent servir également à évaluer l'efficacité de campagnes publicitaires, à livrer des offres plus appropriées et à ajuster le contenu de sites Web. Les Web bugs ne sont donc rien d'autre qu'un nouvel outil d'écoute rapportant une partie de nos faits et gestes pour analyser notre comportement au bénéfice (commercial et financier) de sociétés de marketing et autres annonceurs qui essayent de consolider des statistiques sur les visiteurs communs à leurs sites affiliés ou d'alimenter des fichiers clients (ce qui représente d'appréciables sources de revenus). Du côté des sites, les informations recueillies sur les internautes sont utilisées pour proposer aux annonceurs des espaces publicitaires mieux ciblés. Seulement cet outil est plus vicieux dans le sens où ils ne laissent aucune trace. Pour les internautes, le danger provient du fait que les Web bugs peuvent être associés à des cookies et accéder à ceux existants préalablement sur le disque dur à condition que ces cookies proviennent du même site Web ou de la même compagnie publicitaire. Néanmoins, les Web bugs peuvent avoir une utilité non négligeable pour les administrateurs de sites Web en les aidant à mieux contrôler le contenu de leurs pages. En effet, les pixels espions leur donnent la capacité d'établir des statistiques précises sur la fréquentation des pages de leur site et par conséquent, de mieux connaître leur audience. C'est donc aussi pour cette raison que les Web bugs sont de plus en plus utilisés aujourd'hui.

Un certain nombre d'internautes savent qu'ils sont profilés quand ils voient une bannière publicitaire, mais ils ne peuvent pas voir les Web bugs[41]. C'est la raison pour laquelle ils sont employés pour tracer les internautes dans des pages où les annonces publicitaires ne sont pas présentes ou à l'intérieur de sites dont les visiteurs ne s'attendent pas à être profilés. Richard Smith, expert en sécurité informatique, affirme qu'une large variété de sites contiennent des Web bugs, allant de sites médicaux très sérieux jusqu'aux sites pornographiques. Il a mis au point sur son site Web (http://www.tiac.net/users/smiths) un petit

[40] A chaque fois que la page sera appelée par un internaute, un compteur sera incrémenté chez le certificateur d'audience permettant ainsi au Webmaster du site d'estimer son audience. Il faut noter que, selon les systèmes, des informations plus ou moins précises sont véhiculées par le certificateur. Ce système est assez largement utilisé (avec ou sans l'emploi de cookies) sur des sites persos, associatifs, indépendants, non marchands mais surtout sur des sites commerciaux (c'est même pour certains, le « coeur » de leur métier... et de leur chiffre d'affaires).
Richard Smith, responsable technologique à la Privacy Foundation, fut un des premiers, en novembre 1999, à alerter l'opinion publique des dangers pour la vie privée que représentait cette nouvelle menace. A l'origine, les Web bugs étaient principalement utilisés pour mesurer l'impact des publicités envoyées par e-mail. En fait, selon Smith, ce procédé était déjà utilisé couramment par des sociétés publicitaires pour traquer et mesurer l'audience des bannières publicitaires insérées dans des courriers électroniques au format HTML envoyés en plusieurs millions d'exemplaires. Les pixels espions, au moyen de cette pratique qui a encore lieu aujourd'hui, permettent à ces sociétés de déterminer combien de personnes ont lu leurs lettres, le nombre de fois que ces personnes les ont lues, et si elles les ont transmises à d'autres personnes.

[41] Il existe néanmoins un logiciel gratuit, nommé Bugnosis (voir chapitre suivant), qui permet de les détecter, sans pourtant les bloquer. Une solution tout aussi efficace serait de regarder le code source de la page HTML, seulement c'est une opération pénible et réservée aux internautes expérimentés sachant interpréter le langage HTML.

moteur qui recherche les Web bugs. Une petite recherche sur ce site indique qu'il y a plus de 80 000 sites différents qui contiennent des Web bugs en provenance de DoubleClick, une grande régie publicitaire en ligne.

Employé seul, le pouvoir de nuisance maximal d'un Web bug est en fait atteint lorsqu'il est inséré dans un courrier électronique écrit en HTML[42], que ce mail soit lu sur un site de Webmail[43] ou avec un logiciel spécifique installé sur la machine de l'internaute. C'est de cette façon que les Web Bugs permettent d'outrepasser les limites propres aux cookies en parvenant à associer l'adresse e-mail d'un utilisateur à un cookie[44]. Cependant, à elles seules, ces « images-pixels » sont incapables de puiser dans un carnet d'adresses, d'envoyer des copies d'un courrier électronique ou de scanner les fichiers sur un disque dur. Pour réaliser ce

---

[42] Ceci permet d'établir une association entre l'adresse e-mail et l'adresse IP (utile surtout si l'adresse IP est statique) ou un cookie (plus intéressant car stable, quasi permanent et contenant très souvent un GUI), et le site malveillant pourra mettre un nom (si l'adresse e-mail mentionne le nom de son propriétaire) sur un comportement de navigation et, surtout, valider le courrier électronique du visiteur (qui pourra ensuite faire l'objet d'un spamming abusif). De plus, l'utilisation des Web Bugs permet de connaître non seulement l'identité de la personne recevant le message piégé mais aussi à quel moment elle l'a lu. Mais ce n'est pas tout, si le message est reçu par un logiciel de messagerie compatible HTML ou un site de Webmail, alors tout le parcours d'un e-mail contenant le message original peut être connu par l'émetteur d'origine. Autrement dit, si nous faisons suivre un message ainsi piégé, la personne destinataire de ce nouveau message sera également « connue » de l'expéditeur initial. Tous les destinataires suivants sont donc susceptibles d'être identifiés si le code d'origine est resté suffisamment intact. Si un correspondant désactive l'option HTML dans son logiciel de messagerie, le mouchard sera à nouveau opérationnel dès qu'il sera reçu par une autre personne qui n'aura pas effectué cette manipulation. De plus, avec le développement des technologies du Web dans d'autres domaines, les Web Bugs s'étendent désormais aux documents bureautiques comme dans tout document issu d'une application acceptant le code HTML. Lorsqu'ils sont intégrés dans un tel document, conçu avec MS Office par exemple, ils permettent d'établir des statistiques sur son utilisation (en connaissant, par exemple, le nombre de fois qu'une personne l'a compulsé), de le suivre à la trace dans toutes ses pérégrinations et de recueillir des informations sur ceux qui les consultent, à condition que ces personnes soient connectées à Internet en permanence. La possibilité d'insérer des Web Bugs dans des documents Word, PowerPoint ou Excel augmente encore les risques de développement des pratiques de profilage peu plaisantes.

[43] C'est un système de messagerie qui utilise des pages Web en guise d'interface. Un utilisateur peut donc y avoir accès de n'importe où. Un Webmail est normalement gratuit, mais pour obtenir un compte, l'utilisateur est souvent amené à communiquer des informations personnelles au site en question. De plus, les messages étant envoyés sur une page Web classique, cela permet au fournisseur de ce service e-mail d'inclure (en dehors du message lui-même) des publicités personnalisées sur la page HTML où s'affiche le message. C'est pourquoi le Webmail est très souvent sponsorisé et de nombreuses bannières publicitaires sont affichées.

[44] La technique est relativement simple. Lorsqu'une compagnie établit un mailing, elle dispose nécessairement d'un fichier d'adresses e-mail et peut personnaliser les messages à envoyer. La particularité de ce mailing est son format : l'HTML. Dans le cas qui nous concerne, il s'avère aisé d'insérer automatiquement une image, associée à un script (c'est lui qui déclenchera la collecte d'informations), dans tous les mails en spécifiant, en paramètre d'URL, l'adresse e-mail du destinataire. Ce qui donne (on reconnaît là un hyperlien invisible) :

```
<img width=1 height=1 border=0 src="http://www.serveur_à_cookie.com/image.gif?e-mail=user@adresse.com">
```

Dans ces conditions, dès que l'utilisateur ouvre le message (la plupart des logiciels de messagerie électronique sur le marché sont à l'heure actuelle compatible avec le format HTML, que ce soit Outlook Express, Netscape Messenger ou même Eudora), une requête HTTP GET, générée par la messagerie, est envoyée au serveur pour récupérer l'image GIF en question. Bien entendu, cette requête contient plusieurs informations dont deux qui sont plus particulièrement intéressantes : le numéro d'identification unique du cookie et l'adresse e-mail de l'utilisateur. Il suffit alors de retrouver ce numéro dans une base de données client et lui associer l'adresse e-mail correspondante. Ensuite, lorsque cet internaute surfera sur des sites dont les cookies proviennent du serveur qui stockait l'image GIF du message initial, la personne sera clairement identifiée et surveillée. C'est simplement cela un Web bug ! Mais cette technique s'avère redoutablement efficace lorsque le fichier client atteint plusieurs millions d'adresses.

genre de malfaisances, il faut recourir à des scripts ou à des exécutables comme on pourrait en trouver dans certains codes néfastes situés au sein d'applets Java, de JavaScripts ou de composants ActiveX. L'utilisation des Web bugs a tout de même fait l'objet de nombreuses critiques de la part de groupes de défense des libertés individuelles pour lesquels ils représentent une nouvelle menace. En fait, les Web Bugs ne cachent aucune innovation technologique (ce sont des hyperliens invisibles associés à un cookie et/ou des scripts, ceux-ci collectant les données relatives à l'internaute), tout au plus un nouveau mouchard informatique... Mais il faut rester vigilant car les Web bugs sont beaucoup plus répandus sur les sites Internet qu'on pourrait le penser[45]. Qui plus est, désactiver les cookies dans son navigateur ne suffit pas, hélas, pour s'en protéger, et aucune parade efficace n'a pour l'instant été trouvée (sauf par certains logiciels ou services de protection de la vie privée, cfr. Chapitre suivant pour les connaître).

I.7. La mémoire cache.

Tout d'abord, faisons une petite constatation pour bien comprendre l'enjeu de cette technique : lorsque nous sommes allés sur Internet, chacune des pages que nous avons consultée a sans doute laissé des traces sur notre propre ordinateur. En effet, notre navigateur crée, lors de son installation, un répertoire qui va servir de mémoire cache (ou antémémoire, mémoire tampon). L'utilisation d'une telle mémoire n'est rien d'autre qu'un procédé pour améliorer le temps de chargement des pages que nous visitons souvent, tout en réduisant la charge du réseau. Cette mémoire cache est présente sur la plupart des navigateurs, et donc cette pratique est souvent utilisée. Bien entendu, elle peut être désactivée si nous ne souhaitons pas en profiter (remarquons qu'elle est toujours activée dans la configuration par défaut des browsers les plus répandus). Rappelons aussi que chaque cookie posé par un serveur est également une trace qui reste présente sur notre disque dur et indique notre passage sur ce serveur.

Mais comment ça marche ? Lorsque nous lançons une requête au moyen de notre navigateur, celui-ci va d'abord chercher dans le répertoire de notre système qui lui sert d'antémémoire si la page Web que nous avons demandée ne s'y trouve pas déjà. Si elle s'y trouve, c'est qu'elle a déjà été visionnée auparavant (et donc que nous l'avons téléchargée précédemment à partir du serveur hébergeant le site dont cette page fait partie). Si elle ne s'y trouve pas, notre navigateur effectuera alors la requête correspondante sur le réseau, comme nous l'avons vu ci-dessus. Une fois que la réponse est parvenue au navigateur, il l'enregistre dans ce répertoire tout en l'affichant à l'écran. Ainsi, la prochaine fois, si la même demande est à nouveau effectuée, il ira simplement lire la page qu'il a stockée antérieurement sur notre disque dur. Nous constaterons alors que le résultat s'affiche plus vite que si notre navigateur

[45] Selon une étude réalisée par la société britannique de gestion de sites Cyveillance, l'utilisation des Web bugs, a augmenté de près de 488 % durant les trois dernières années. Comme ils l'indiquent dans leur compte-rendu, les auteurs de l'étude, qui ont comparé plus d'un million de pages Web entre 1998 et 2001, révèlent qu'il y a 5 fois plus de chance de visiter une page Web contenant un Web bug en 2001 qu'en 1998. Il y a quatre ans, 0,7 % des sites utilisaient ces mouchards. Ils seraient 3,9 % à adopter cette pratique en 2001. Les sites personnels sont les premiers à utiliser les Web bugs, 18 % d'entre eux y ont ainsi recours (il y a quatre ans la proportion de pages personnelles infectées n'était que de 0,5 %), mais dans la plupart des cas les mouchards n'ont pas été placés par les auteurs du site, mais par son hébergeur (beaucoup de ces pages personnelles sont hébergées gratuitement et, en échange de cette gratuité, les Webmasters doivent la plupart du temps accepter d'afficher des bannières ou des pixels espions). Plus préoccupant, 16 % des pages d'accueil des 50 plus importants sites de marques contiennent des Web bugs, ce qui n'est pas sans soulever quelques questions quant à l'utilisation des données ainsi recueillies.

avait lancé réellement la requête sur le réseau et attendu la réponse provenant du serveur. Par conséquent, le fonctionnement de cette mémoire cache revient à stocker plus près (sur notre ordinateur) une information (une page Web) qui sera probablement redemandée ultérieurement afin de ne devoir aller la chercher plus loin (sur le serveur l'hébergeant) qu'une seule fois, et donc d'optimiser son temps de chargement.

Les navigateurs récents permettent de modifier certains paramètres régissant l'activité de cette mémoire cache. De cette façon, nous pouvons notamment indiquer combien d'espace disque elle peut utiliser au maximum ou encore la période après laquelle le navigateur doit vérifier sur le serveur si la page en question a été modifiée. Mais nous pouvons aussi modifier l'endroit où se trouve le répertoire qui stocke toutes les informations que le navigateur reçoit en changeant sa position par défaut, voire même annuler totalement l'emploi de cette mémoire cache ou supprimer tous les fichiers se trouvant dans son répertoire (il y a toujours moyen d'effacer ces fichiers par nous-même au cas où).

Il faut savoir qu'il existe en pratique deux utilisations possibles de la mémoire cache : il y a la mémoire cache utilisée par notre navigateur, elle est donc stockée sur notre propre ordinateur ; ensuite, il y a la mémoire cache utilisée par un serveur proxy, qui n'est autre qu'un serveur intermédiaire employé pour accélérer la navigation de ces multiples utilisateurs grâce à une vaste antémémoire rapide et partagée (on en trouve habituellement chez les fournisseurs d'accès et dans les grandes organisations). Dans ce cas, la mémoire cache est contrôlée par l'ordinateur qui héberge le serveur proxy en question (par conséquent, son contenu est stocké sur le disque dur de ce même ordinateur).

En enregistrant les pages consultées dans un répertoire situé sur notre ordinateur, le navigateur laisse une trace de notre passage. Assurément, cette trace est sur notre disque dur et donc pas sur celui d'un serveur extérieur. D'ailleurs, ceux-ci ignorent totalement si nous en faisons usage. Néanmoins, il faut savoir que si nous pensions consulter un site Web en toute discrétion sur un ordinateur autre que le nôtre (par exemple celui que nous utilisons sur notre lieu de travail ou dans un « cyber-café »), nous pourrions être surpris. En effet, si la mémoire cache est opérationnelle, il suffit à quelqu'un de venir après notre passage, sur la machine que nous avons utilisée, pour ouvrir le répertoire qui contient les fichiers de l'antémémoire, et reconstruire entièrement notre parcours en les classant par ordre chronologique ou même examiner les pages que nous avez visitées durant notre session. On en déduit donc qu'il n'y a donc aucune raison de se méfier de l'activité de la mémoire cache dès lors que nous avons pris soin d'en effacer le contenu avant de quitter la machine sauf bien sûr si nous ne redoutons pas que les traces de notre passage restent présentes ou même qu'elles puissent être consultées après nous[46] !

Sur un serveur proxy, cela peut poser un problème plus délicat à résoudre dans la mesure où la possibilité de conserver la trace des requêtes effectuées par ses utilisateurs est dans la main de son administrateur. Bien sûr, le fait que nos requêtes passent par le serveur proxy peut nous échapper et donc nous n'avons aucun moyen d'empêcher la conservation éventuelle de ces fichiers. C'est donc nous qui devons décider, si on en a le choix, si nos requêtes doivent transiter par tel ou tel serveur proxy.

[46] Sur Netscape, même après avoir vidé la mémoire cache et supprimer l'historique des liens visités, le navigateur peut encore révéler beaucoup de choses. En effet, il suffit d'indiquer dans la barre d'adresses de ce navigateur la commande *about:global* ou encore *about:cache*, au lieu d'une URL habituelle, pour faire apparaître des informations surprenantes que l'on avait pourtant cru avoir effacées pour de bon !

I.8. Cas concret d'utilisation d'une méthode de profilage

A la lumière de ce que nous avons vu précédemment, nous allons examiner un cas concret : comment les compagnies de bannières publicitaires interceptent les mots clés que nous tapons sur les moteurs de recherche et utilisent les cookies pour pister nos mouvements à travers de nombreux sites. Nous allons voir que ces compagnies, bien qu'étant des sites tiers vis-à-vis de notre communication avec un site déterminé (un moteur de recherche, dans l'exemple que nous examinerons) sont néanmoins capables de capter certains éléments clés de cette communication (les mots clés que nous avons tapé) afin d'injecter en temps réel une bannière publicitaire ciblée sur la page Web du site en question. Nous verrons aussi que les différentes techniques exposées ci-dessus ne sont absolument pas exclusives l'une par rapport aux autres. L'exemple exposé[47] est largement inspiré de celui présenté dans l'article « Les traitements invisibles sur Internet » de J.-M. Dinant (cfr. [1], pages 10 à 14). Il s'agit d'une technique de cybermarketing utilisée par la société américaine DoubleClick (http://www.doubleclick.net) qui est un des plus grands leaders du marché des annonceurs en ligne et qui se vante de fournir plusieurs millions de bandeaux publicitaires ciblés par jour. Cette société, spécialiste américain du bandeau publicitaire, a constitué une base de données dressant des profils d'utilisateurs à qui on envoie une publicité choisie parmi un stock d'environ 50 000.

Lorsqu'un internaute se connecte au site d'un moteur de recherche (Altavista ou Lycos), la requête HTTP envoyée par son navigateur usuel communique au serveur de ce site les données suivantes : adresse TCP/IP de l'internaute (pour envoyer la réponse), le type de navigateur (modèle, version linguistique,...) et le système d'exploitation, utilisés par ce dernier ainsi que la (ou les) langue(s) acceptée(s). Ce serveur répondra en envoyant le code HTML de la page d'accueil du site demandé (s'il l'héberge, bien entendu). Ensuite, le navigateur commence à afficher le contenu de la page qu'il reçoit. Admettons que cette page contienne un hyperlien invisible vers le site de DoubleClick. Le navigateur envoie alors une nouvelle requête à destination du site de DoubleClick et lui transmet les données suivantes : adresse TCP/IP de l'internaute, le type de navigateur et le système d'exploitation, utilisés par ce dernier ainsi que la (ou les) langue(s) qu'il accepte, la page référente (à savoir l'adresse de la page en cours de chargement) et les cookies préalablement enregistrés par DoubleClick (si l'internaute en avait déjà sur son disque dur). Si l'internaute a activé la fonctionnalité d'avertissement lors de la réception de cookies sur son navigateur, il pourra constater que la société DoubleClick veut déposer un nouveau cookie sur sa machine avec une date d'expiration éloignée de plusieurs dizaines d'années. Ce cookie contient un numéro d'identification unique et permanent affecté par DoubleClick à l'utilisateur[48]. Supposons qu'il accepte ce cookie. Après quelque temps, l'affichage de la page sera complet et comprendra une bannière de pub issue du site de DoubleClick.

Imaginons que l'internaute effectue une recherche en tapant certains mots clés dans la zone de texte prévue et en cliquant sur le bouton « Rechercher ». Le navigateur envoie une nouvelle requête HTTP au serveur hébergeant le moteur de recherche ; cette requête indique l'URL recherchée sur le site, à savoir celle correspondant à une recherche des mots clés indiqués (la requête indiquera explicitement les mots clés recherchés, qui sont écrits en toute

[47] Pour un autre exemple « en direct » et s'appuyant sur les mêmes techniques, cfr. [27].

[48] Ce cookie permettra de déterminer avec certitude combien de fois une bannière a été vue par un utilisateur donné mais aussi de compléter et affiner les données de son profil au fur et à mesure de ses sessions.

lettre dans l'URL). Comme le code HTML de la réponse du moteur de recherche inclut une bannière (souvent une image GIF animée) provenant du site de DoubleClick, le navigateur effectue une requête implicite vers le serveur de ce site avec comme page référente l'adresse de la page en cours d'affichage (le champ HTTP_REFERER indiquera donc l'URL contenant les mots clés de la recherche effectuée par l'internaute). Cela montre, qu'avant même de renvoyer sa bannière publicitaire, le site de DoubleClick est mis au courant précisément de la recherche en cours. Connaissant les mots clés, il a la possibilité d'expédier au navigateur une première bannière réellement susceptible d'intéresser l'internaute, bannière choisie parmi les milliers qu'il a en stock. Celle-ci sera d'ailleurs affichée en même temps que les résultats de la recherche (tout ce processus s'opère en guère plus de 20 millisecondes). De plus, le programme de navigation indiquera à nouveau qu'un cookie de DoubleClick a été envoyé. Celui-ci a pour but d'affiner le profil de l'internaute en contenant une nouvelle valeur codée.

Imaginons encore que l'internaute décide de cliquer sur la publicité qui lui est présentée. Si on regarde de plus près l'hyperlien lié à cette bannière, on pourrait constater que l'activation de celui-ci aurait pour effet de l'envoyer sur le site de DoubleClick (l'URL contenant explicitement le nom DNS d'un serveur de DoubleClick). Or il n'en est rien, l'expérience montre en effet que le clic amène sur une page d'un site appartenant à la société dont la pub vend les services ou les produits (en d'autres mots, l'annonceur). Il n'y a rien d'extraordinaire à cela, puisque ce n'est rien d'autre que la manifestation d'une redirection automatique (caractéristique cachée du fonctionnement des navigateurs). L'avantage de DoubleClick dans cette façon d'opérer est d'être averti que l'internaute a porté son attention pour cette bannière spécifique (c'est le « successful targeted click-through »). Cela permet à nouveau de perfectionner son profil et aura sans doute pour effet de lui proposer ultérieurement des produits de la même entreprise ou du même genre. En moyenne, ces sociétés estiment obtenir un taux de réussite de 40%. Une fois sur le site de l'annonceur, les prestataires continuent à analyser la navigation des utilisateurs ; ils recensent notamment qui achète, qui abandonne le site et à quel endroit précisément cela se produit.

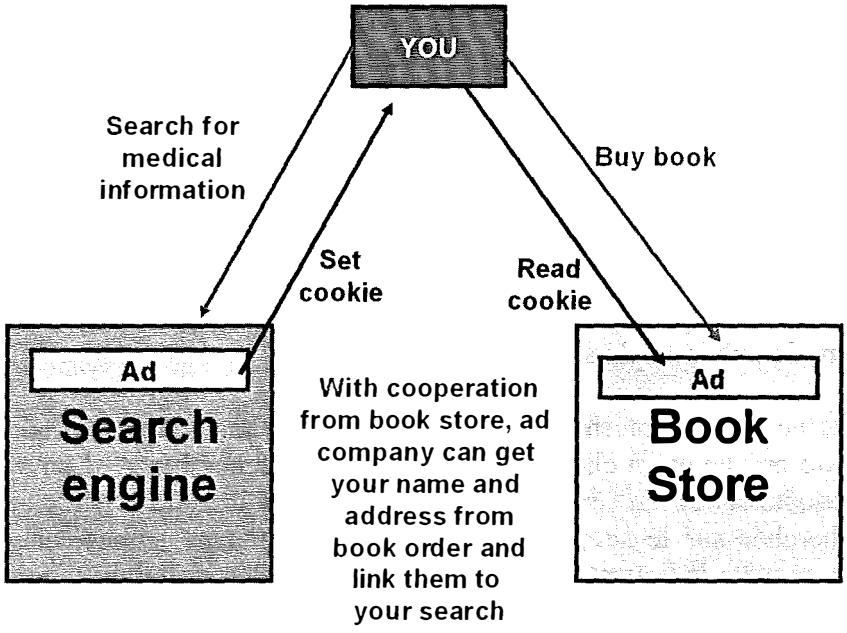

Figure I-1 (source : référence [13] page 7)

Remarquons aussi que la firme de cybermarketing ne proposera jamais que des publicités recommandant les produits de ces propres clients. Un internaute peut donc se trouver dans une situation où, utilisant un moteur de recherche pour trouver des informations sur un produit d'une certaine marque, se voit proposer des produits du même type mais d'une marque concurrente. Conscients des discussions en cours autour de la protection de la vie privée et des risques de contestation auxquels ils pourraient avoir à faire face, les prestataires en question certifient ne pas conserver dans leurs fichiers l'identité patronymique des utilisateurs, pas plus que leur adresse e-mail. DoubleClick a ainsi été amené à mettre en place sur son propre site une procédure d'opt-out qui inhibe la lecture du cookie par ses serveurs. En même temps, cela permet d'éliminer une catégorie d'utilisateurs que l'on pourrait qualifier de publiphobes qui de toute façon est peu réactive et avec laquelle on a peu de chances de réaliser des achats « on line ». DoubleClick déclare enregistrer quotidiennement entre 5 et 10 applications de la procédure d'opt-out.

Ainsi, grâce à la coopération des navigateurs, les sociétés de marketing sur Internet commercialisant les espaces publicitaires (Rep-Agency[49]), dont DoubleClick n'est qu'un exemple, peuvent connaître, pour les centaines de pages des milliers de sites supports, le chemin de chaque internaute à travers ces pages. Comme on l'a vu, l'origine de cette capacité se trouve dans les navigateurs qui communiquent de nombreuses données aux serveurs à l'insu de l'internaute, qui ne peut effectivement s'y opposer (mais, s'il est vigilant, il peut tout de même parvenir à déjouer dans une certaine mesure ce vicieux stratagème, comme nous le verrons au cours du deuxième chapitre).

Les firmes de cybermarketing peuvent rassembler l'ensemble des parcours des utilisateurs des sites supports en collectant leurs adresses IP, leurs anciens parcours (grâce aux cookies), les actions effectuées sur les sites supports (par exemple, les mots clés recherchés sur les moteurs de recherche affiliés), les publicités déjà envoyées (et donc visionnées) et les publicités ayant suscitées un intérêt positif. Toutes ces données sont datées et sont totalement indépendantes des adresses IP variées que la même machine aurait pu recevoir au cours de ces diverses visites sur les sites supports. Ainsi les sociétés de cybermarketing peuvent recueillir discrètement, sur toute la planète, les agissements précis et détaillés de chaque internaute en particulier.

D'après les agences de publicité en ligne, les cookies et les autres techniques de profilage ont pour unique but d'aider les consommateurs et les sites Web. Elles affirment, sous le feu des nombreuses accusations dont elles font l'objet, que les informations collectées sont gardées privées et sont leur seule propriété. Mais l'inquiétude de nombreuses associations de défense des consommateurs a monté d'un cran lorsque DoubleClick a racheté Abacus Direct en 1999, puisque Abacus dispose d'une base de données qui contient les profils de consommation détaillés de plus de 90% des ménages américains. La fusion de la base de données de DoubleClick contenant les profils de millions d'internautes avec les données personnellement identifiables d'Abacus a fait explosé une tempête de critiques, et a mené à une enquête du gouvernement. DoubleClick a alors fait marche arrière en abandonnant ses plans d'unification des deux bases de données, mais jusqu'à quand ? On comprend que le problème prenne une ampleur plus importante lorsqu'une compagnie peut effectuer un rapport direct entre le GUI d'un cookie et des données personnelles identifiantes telles qu'une adresse

[49] Leur rôle consiste à offrir aux annonceurs une prestation globale intégrée en matière de campagne publicitaire sur Internet : media planning, ciblage de type « *one-to-one* », contrôle, suivi de l'impact et reporting client. Ces annonceurs peuvent n'avoir aucun rapport avec les sites sur lesquels leur bannière va être affichée.

ou un numéro de téléphone. DoubleClick a d'ailleurs fait l'objet de plusieurs plaintes et a été poursuivie par le procureur général de l'état du Michigan parce qu'elle avait violé les lois de protection du consommateur en n'informant pas les internautes qu'elle plaçait régulièrement des cookies sur leurs disques durs, ni de l'usage qu'elle en faisait. La société s'est défendue en se déclarant contractuellement obligée de conserver, pour elle seule, ces informations et ces outils (c'était, selon elle, le seul moyen de perfectionner la publicité sur Internet). Depuis, de nombreuses voix se sont élevées pour protester contre l'utilisation des données personnelles, notamment nominatives, et des croisements de fichiers, suite à l'affaire Abacus.

I.9. Conclusion

Comme nous l'avons constaté, la combinaison des différentes techniques que sont le bavardage des navigateurs, les hyperliens invisibles, la redirection automatique et les cookies permet d'envoyer en temps réel des publicités ciblées à tout internaute qui emploie un navigateur installé par défaut en ne connaissant que son adresse IP (il n'est pas nécessaire de connaître son nom ou son adresse postale, par exemple). Dans ce type de marketing, les cookies ne sont pas vraiment essentiels car ils ne servent qu'à maintenir le profilage en mémoire entre deux connexions du même internaute aux différents sites supports. Seule l'adresse IP est indispensable pour assurer le profilage au cours d'une connexion particulière. Néanmoins, il ne faut pas négliger l'importance des cookies car sans eux le profil de l'internaute qui ne possède pas une adresse IP permanente ne pourrait pas être affiné sans cesse connexion après connexion. Ce profilage n'est pas en soi inhérent au protocole HTTP tel que défini par le World Wide Web Consortium (http://www.w3.org). En outre, la définition du protocole HTTP 1.1 a explicitement attiré l'attention de l'industrie sur les risques d'atteinte au respect de la vie privée lors de son implémentation[50].

N'oublions pas aussi que si théoriquement toutes ces techniques (qui rendent possible le polymorphisme des sites Web) sont uniquement utilisées dans un but commercial, rien n'empêche qu'elles soient aussi employées pour d'autres desseins (censure, discrimination…) et ce sans qu'à aucun moment, l'internaute ne puisse connaître les informations contenues dans son profil, les modifier ou même s'opposer à leur usage.

Que dire de l'envoi d'informations dans des formulaires en ligne ? Bien entendu, si nous choisissons volontairement de divulguer des informations au site, notamment en renvoyant un formulaire qu'il nous présente, nous fournissons sciemment au site des informations personnelles, celles que nous avons entrées. Le site est alors libre d'enregistrer ces informations dans sa base de données et de les utiliser à sa guise. Pour notre protection, de nombreux sites établissent désormais volontairement des politiques de confidentialité qui régissent l'utilisation des informations transmises, dictant ce qu'ils feront et ne feront pas avec les données que nous leur fournissons (pour plus de détails sur ce genre de politiques, voir chapitre II). Chaque site détermine sa propre politique de confidentialité et la rend disponible pour nous permettre de la consulter librement.
N'oublions pas qu'il n'existe aucun contrôle des sites sur le bon respect de leur charte de confidentialité et qu'ils peuvent affirmer ce qu'ils veulent. Ainsi, en fin de compte, la décision finale concernant la transmission volontaire d'informations à un site dépend de la confiance

[50] Le mot « vie privée » est mentionné 18 fois dans la RFC 2068, http://www.w3.org/Protocols/rfc2068/rfc2068 (cfr. [21]), et 21 fois dans la RFC 21 fois dans la RFC 2616, http://www.w3.org/Protocols/rfc2616/rfc2616 (cfr. [23]).

que nous lui accordons. Nous pouvons être enclin à croire les déclarations de la politique de confidentialité de *http://www.dignedeconfiance.com* alors que nous pourrions avoir des raisons de douter de celle de *http://www.sitedouteux.com*

Nous entrons fréquemment les mêmes informations dans les formulaires de différents sites. Par exemple, tous les sites d'achat nous demanderont probablement notre nom, l'adresse de livraison et notre numéro de carte de crédit. Il est fastidieux de toujours devoir entrer ces données. Nous pouvons alors demander au gestionnaire de formulaires de notre navigateur (si celui-ci en incorpore un, à moins d'utiliser un logiciel spécifique) de les enregistrer et de les entrer automatiquement dans les formulaires futurs. Le gestionnaire enregistre les informations sur la machine locale et non sur un site Web distant. Lorsqu'il entre les informations enregistrées dans un formulaire, celles-ci ne seront pas envoyées au site tant que nous n'aurons pas soumis le formulaire. Nous sommes de nouveau en position de contrôle, aucune information n'étant fournie sans notre autorisation, à condition toujours d'avoir une certaine confiance vis-à-vis des sites auxquels nous acceptons de fournir ces informations.

Quelques mots à propos de la divulgation des mots de passe : un grand nombre d'internautes sont abonnés aux services de différents sites. Ce genre d'abonnement consiste la plupart du temps en une sélection d'un nom d'utilisateur et d'un mot de passe. Chaque fois que nous revisitons ce site, nous remplissons et envoyons un genre de formulaire contenant le nom d'utilisateur et le mot de passe choisis. Pour éviter de devoir mémoriser un mot de passe différent pour chaque site, en particulier ceux que nous ne visitons pas souvent, nous avons peut-être utilisé le même mot de passe partout. Il en va de même pour le nom d'utilisateur, à moins qu'il ait déjà été choisi par quelqu'un d'autre. Chaque site auprès duquel nous sommes enregistrés dispose donc de deux informations importantes, notre nom d'utilisateur et notre mot de passe. Si nous utilisons toujours les mêmes, un administrateur de site peu scrupuleux dispose d'informations suffisantes pour prendre notre place auprès des site auxquels nous sommes déjà abonnés. Nous pouvons ne pas nous en préoccuper si cela ne nous pose pas de problème que quelqu'un se connecte au même site d'actualités et consulte les dernières nouvelles mondiales. C'est toutefois plus délicat si quelqu'un réussit à connaître, par exemple, le nom de notre courtier, à s'y connecter en utilisant notre identité et à effectuer des transactions boursières.
Pour nous protéger, nous pouvons évidemment utiliser un mot de passe différent sur chaque site. Cela implique toutefois que nous devons nous en souvenir. Le gestionnaire de mots de passe de notre navigateur (si celui-ci en est doté, à moins d'employer un programme spécifique) peut nous aider en mémorisant les noms d'utilisateur et mots de passe employés lors de notre dernière connexion à un site et en saisissant automatiquement ces informations dans le formulaire de connexion lors de notre prochaine visite. Nous pouvons envoyer le formulaire de connexion avec ces valeurs saisies ou les modifier avant envoi si elles sont inexactes.
Un tel gestionnaire de mots de passe nous permet également d'afficher les noms d'utilisateur stockés pour différents sites. Il permet aussi de les supprimer de manière sélective. Ainsi, il permet de simplifier l'utilisation des services offerts par de nombreux sites Web.

Lors du téléchargement d'un fichier par le protocole FTP, il est d'usage, par courtoisie, de communiquer son adresse électronique au site. Il importe donc de jeter un œil attentif lors de la configuration de son browser si on ne souhaite pas qu'une telle information soit divulguée sans en être averti (on peut savoir que notre navigateur utilise le protocole FTP pour la demande d'un fichier puisque son URL commence par « ftp:// » et non par « http:// »). Dans de nombreux navigateurs, l'un des paramètres de configuration détermine si l'adresse e-

mail doit être ou non envoyée comme mot de passe lors du téléchargement de fichiers. Heureusement, cette option est le plus souvent initialement configurée pour ne pas envoyer d'adresse électronique ; celle-ci ne sera alors divulguée que si l'on modifie ce paramètre.

Doit-on masquer son adresse IP ? Lorsque nous demandons à afficher une page d'un site, le navigateur doit indiquer au site notre adresse IP pour qu'il sache où envoyer la page demandée. Il s'agit en quelque sorte de notre adresse de retour. Notre fournisseur d'accès à Internet possède de nombreuses adresses IP et il en choisit une pour nous à chaque session. Nous en recevons donc une nouvelle à chaque connexion avec notre fournisseur. Bien sûr, comme évoqué ci-dessus, certains utilisateurs ont une adresse IP fixe utilisée à chaque connexion à Internet. Ces utilisateurs sont toutefois minoritaires (mais leur nombre augmente de plus en plus suite à l'engouement pour l'Internet haut débit) mais, si nous en faisons partie, nous le savons incontestablement. Ainsi, à moins d'avoir été informé du contraire, nous pouvons supposer que nous recevons une nouvelle adresse IP à chaque session.

Bien que cette adresse soit temporaire, nous ne souhaitons peut-être pas que ces informations soient transmises au site que nous visitons. Mais si le navigateur ne fournit pas ces informations, le site ne saura pas où envoyer la page demandée. Ces informations doivent donc être absolument transmises par le navigateur. Comment faire dans ce cas où nous ne souhaitons absolument pas révéler notre adresse IP au site visité ? Nous devons utiliser un serveur intermédiaire digne de confiance. Nous devons lui indiquer l'URL du site dont nous souhaitons consulter les pages. Le site intermédiaire demande alors les pages pour nous, en utilisant sa propre adresse IP comme destination. Lorsqu'il reçoit ces pages, il nous les transmet. Le site qui a possède les pages en question n'a donc jamais accès à notre adresse IP. Nous reparlerons de ces techniques au cours du deuxième chapitre de ce mémoire.

Internet a été conçu comme un réseau ouvert au niveau mondial, au travers duquel des informations peuvent être partagées. Il est cependant nécessaire de trouver un juste équilibre entre la liberté d'Internet et la protection des données personnelle de ces utilisateurs.
Des quantités énormes de données sont récoltées sur ce réseau à propos de ces utilisateurs, alors que souvent ces derniers n'en sont même pas conscients. Ce manque de transparence vis-à-vis d'eux doit être corrigée de manière à atteindre un niveau suffisant de protection des consommateurs et de leurs données personnelles.
Les protocoles sont des moyens techniques qui déterminent en fait comment les données sont récoltées et traitées. Les navigateurs et les logiciels jouent également un rôle important en acceptant notamment qu'un identificateur permette de relier l'internaute à ses activités sur le Internet. Il en va donc de la responsabilité de ceux qui conçoivent et développent ces produits, de fournir à leurs futurs utilisateurs des produits respectant la vie privée.

Quelle attitude adopter ? Tout d'abord, il faut rester réaliste. Sur Internet comme ailleurs, l'absence de traces nécessite des efforts, mais la surveillance aussi. Il ne faut pas sombrer dans la paranoïa ni dans l'utopie. Internet n'est pas un monde virtuel. C'est un réseau informatique. A ce titre, lorsqu'on l'utilise, on laisse des traces. On peut être observé, il faut en être conscient. Cela ne signifie pas forcément que l'on est effectivement observé. Il convient de rester vigilant, d'apprendre à mesurer les risques et à agir en conséquence.
Il importe aussi de s'informer. Dans la vie quotidienne, nous prenons, sans même y penser, plusieurs précautions pour protéger notre vie privée. Elles sont naturelles. Sur Internet, tout est nouveau. Les outils évoluent tellement vite que l'on a parfois le sentiment d'être dépassé. Il faut déjà beaucoup de temps pour apprendre à exploiter efficacement les logiciels, en reste-t-il pour s'informer des risques qu'ils font courir à notre vie privée ? Une bonne utilisation

d'Internet suppose que nous ayons un comportement responsable, c'est-à-dire que nous soyons relativement conscients des conséquences de nos actes, en particulier de leurs répercussions sur l'intégrité de notre vie privée. N'oublions pas qu'un internaute est en droit d'être informé de toute collecte de données le concernant, et être en mesure de s'y opposer.

Cela met toujours mal à l'aise de savoir que de tels systèmes de profilage sont mis en oeuvre sans en informer les utilisateurs. Pourquoi les sites les utilisant ne sont-ils pas plus explicites ? L'emploi quasi systématique de ces systèmes ne doit pas banaliser le fait que les internautes sont analysés, « statistisés », comptabilisés, à leur insu. La méfiance légitime qu'on peut avoir envers des entreprises dont le métier est de vendre des données marketing et qui, par le passé, n'ont accepté de donner un peu de transparence à leurs activités que sous la pression des associations de défense de la vie privée, devrait conduire les Webmasters indépendants à avoir un regard plus critique vis-à-vis de ces systèmes.

Quelques conseils pour aider les internautes à se protéger ? Eviter au maximum de donner ses coordonnées réelles dans les formulaires en ligne. Vider la mémoire cache de son navigateur et détruire les cookies à la fin de chaque session. Mais ils peuvent encore mieux se protéger ... s'ils en ressentent la nécessité (cfr. Chapitre II).

Quelles « bonnes pratiques » pour les Webmasters :

- limiter l'usage des cookies au strict minimum (des cookies de session non enregistrés sur le disque dur des internautes sont généralement suffisants pour la plupart des usages)
- éviter de recourir aux Web bugs et aux services des certificateurs d'audience mettant en oeuvre des cookies (mais c'est pas évident lorsqu'il s'agit d'obtenir un hébergement gratuit)
- publier sur son site une « privacy policy » (politique en matière de vie privée ou charte de confidentialité), càd un traité indiquant la nature exacte des moyens mis en oeuvre pour collecter de l'information sur ses visiteurs et l'usage qui est fait de ses informations une fois récoltées, et bien sûr s'y tenir !

Suite à cette présentation, il y a de quoi devenir méfiant vis-à-vis du net. L'internaute a le droit d'être protégé, et doit pouvoir se protéger. Néanmoins, l'apparition de tels « outils » sur Internet constitue un progrès non négligeable dans l'amélioration du confort lors de la navigation (peut-on croire qu'ils aient été développés uniquement dans ce but ?). Malheureusement, l'utilisation de ces techniques faite par les fournisseurs de contenu en ligne a évidemment de quoi nous laisser sceptique. On peut se poser la question de savoir comment des technologies pareilles ont été développées sans avoir au préalable fait l'objet d'une réflexion sur les droits de la vie privée de chacun...

Alors, qui est responsable ? Celui qui est sous la coupe de la justice est le « responsable du traitement ». Mais il s'agit aussi de sensibiliser l'utilisateur pour qu'il fasse valoir ses droits, remplir ses devoirs et faire preuve d'un minimum de prévoyance.
Ne devrait-on pas également sensibiliser davantage les auteurs des outils utilisés sur Internet. Le navigateur est la clé qui nous permet d'entrer sur le Web mais aussi le véhicule qui nous permet de nous déplacer dedans, est-il normal que l'on voyage avec un véhicule qui risque de nous trahir, à notre insu, à la moindre occasion ?

Comment s'assurer que des personnes non autorisées ne puissent pas utiliser des informations nous concernant ? Le meilleur moyen de garder des informations secrètes est

soit de ne pas les confier soit de faire très attention à ceux à qui on les confie. Comme nous l'avons déjà dit, Internet est un réseau public, et il faut prendre en considération que lorsque l'on fournit à quelqu'un son nom, son numéro de téléphone, son adresse ou d'autres informations (à partir d'une page Web, par courrier électronique, ou toute autre méthode), il se peut que cette personne les partagent avec d'autres, sauf si l'on s'est mis d'accord avec le destinataire sur l'utilisation des informations échangées. En fait, fournir son nom, adresse et numéro de téléphone sur le Web, c'est un peu comme être sur une liste dans un annuaire.
Si un site Web demande des informations personnelles, il importe de se renseigner sur l'usage qu'il compte en faire avant de les divulguer. Voici quelques questions que l'on peut se poser avant de révéler ces informations[51] :

- Quels types d'informations à caractère personnel essaie d'obtenir le site ?
- Comment ce site utilisera ces informations et dans quel but ?
- Est ce que ce site partagera les informations obtenues avec d'autres sites, pour des utilisations à caractère publicitaire ?
- Puis-je accéder à une partie ou à la totalité des informations fournies au site, dans un but de vérification et/ou de modification ?
- Comment le site protège-t-il les informations reçues vis-à-vis des hackers ?
- Comment contacter le responsable du site Web si j'ai des questions à poser ou des revendications à signaler ?

[51] Nous conseillons également la lecture des 12 conseils de l'Electronic Frontier Foundation pour protéger sa vie privée en ligne (http://www.bugbrother.com/eff/eff_privacy_top_12.html) ou des lignes directrices pour la protection des personnes à l'égard de la collecte et du traitement de données à caractère personnel sur les « inforoutes » (cfr. [12]).

Chapitre II : Protection de la vie privée sur Internet ? Etude de l'existant et Solutions diverses

Préserver sa vie privée lors de sa navigation sur le Web n'est pas une tâche facile, elle devient un véritable souci pour de plus en plus de personnes, et pas seulement pour celles appartenant aux associations de protection des libertés individuelles. Tout internaute risque de devenir un jour partie intégrante d'une base de données massive qui stockera le profil de ses actions à chaque fois qu'il surfera sur Internet. Malgré les procès intentés contre elles, les compagnies publicitaires, sans législation efficace pour protéger les internautes, donnent libre cours à leurs activités en profilant tous les surfeurs qui croisent leurs annonces et en liant les informations qu'elles collectent à partir de formulaires ou d'enquêtes en ligne aux clickstreams des personnes correspondantes.

Pourquoi vouloir absolument chercher à protéger sa vie privée ? Simplement pour permettre au Netoyen - citoyen du Net - que nous sommes de pouvoir jouir de son droit légitime et légal (au vue de la loi Européenne) au respect de la vie privée et des libertés individuelles. A chacun la liberté de revendiquer ses droits pour qu'Internet reste et demeure une zone de liberté.

La directive communautaire sur la protection des données contient deux principes ayant des conséquences directes sur la conception et l'emploi de nouvelles technologies :

- son principe de « finalité » qui exige que les données à caractère personnel ne soient utilisées que dans un but spécifique et légitime; en d'autres mots, sans une raison légitime, les données à caractère personnel ne peuvent pas être utilisées et l'individu reste anonyme.
- son principe de « sécurité des données » qui exige que les responsables mettent en œuvre des mesures de sécurité appropriées aux risques encourus par les données à caractère personnel stockées ou transmises, en vue de protéger ces données à caractère personnel contre toute destruction accidentelle ou illicite, perte accidentelle, altération, divulgation ou accès, en particulier quand le traitement implique la transmission des données sur un réseau, et contre toute autre forme de traitement illicite.

Le principe de « finalité » susmentionné est le raisonnement sur lequel est basé le concept de technologies de renforcement de la protection de la vie privée (Privacy Enhancing Technologies, ou PET) (cfr. [5, 6]). Ce concept concerne une gamme de technologies qui assurent la protection de la vie privée, notamment en minimisant ou en éliminant la récolte ou le traitement de données identifiables. Les technologies de renforcement de la protection de la vie privée visent ainsi à limiter toute forme de traitement illicite des données à caractère personnel, pour empêcher par exemple les éventuelles destructions, altération et divulgation de ces données. Le but de ce chapitre est, entre autre, de faire un petit tour d'horizon de ces différentes PET.

Nous commencerons ce chapitre par une comparaison des navigateurs les plus répandus qui sont mis à la disposition, gratuitement, des internautes en terme de protection des données à caractère personnel. Nous constaterons que ces différents navigateurs ne sont pas tous égaux de ce point de vue, loin s'en faut... Nous continuerons ce chapitre par une section consacrée aux limites des différentes techniques de collecte de données personnelles abordées dans le chapitre précédent. Ensuite, nous vouerons les deux sections suivantes à un

examen concis des diverses PET, du moins les plus courantes. Signalons que tous les logiciels cités dans ce chapitre sont présentés à titre d'information : il existe tellement de logiciels dont la tâche consiste à améliorer la protection de la vie privée de leurs utilisateurs, et ce dans des catégories très variées, qu'il nous est impossible de tous les présenter (d'ailleurs, il en apparaît de nouveaux très régulièrement) ; par conséquent, notre sélection (cfr. [14, 15, 16, 29]) ne révèle que ceux étant largement répandus, ayant une certaine cote de « popularité » ou nous ayant réellement convaincu de leurs capacités (et de leur facilité d'utilisation) tout en étant gratuit. Il est donc fort probable que l'on puisse trouver sur Internet d'autres logiciels ayant une plus grande efficacité et/ou possédant un plus grand nombre de fonctionnalités. De plus, il existe souvent un équivalent gratuit tout aussi efficace et complet à un logiciel commercial. Il ne faut pas hésiter à utiliser son moteur de recherche favori pour les trouver, et pourquoi pas, les adopter définitivement... Comme la plateforme Windows est certainement la mieux représentée dans la population des internautes, la grande majorité des logiciels référencés sont adaptés à cette plateforme. Néanmoins, il existe presque toujours un programme similaire adapté aux plateformes Apple Macintosh ou Linux.

II.1. Les navigateurs les plus communs

Nous avons identifié 7 risques potentiels d'atteinte à la vie privée d'un utilisateur surfant sur Internet, ceux-ci sont :

- la communication « en clair » de l'adresse IP, au sein des requêtes HTTP dû à l'utilisation des protocoles TCP/IP
- le bavardage des browsers
- les hyperliens invisibles (et les Web bugs)
- la redirection automatique
- les cookies
- le scripting côté client (langages de scripts, applets Java et ActiveX)
- la mémoire cache

Dans cette section, nous examinerons de plus près les différents navigateurs qu'un internaute peut utiliser (nous en avons relevé quatre d'utilisation courante) et nous verrons en quoi leur emploi influe sur les risques auxquels la vie privée de son utilisateur peut être confrontée. Signalons que cette étude a été effectuée entièrement sous une plateforme Microsoft Windows XP Professionnel (équivalent Windows NT 5.1), avec la version adaptée des différents navigateurs.
Voici les 4 navigateurs que nous avons choisi d'étudier (car ils nous semblait les plus représentatifs des choix des internautes) :

- Microsoft Internet Explorer 6.0 (MSIE 6) (version française « belge », car incorporé dans Windows XP configuré d'après les options propres à la Belgique) (http://www.microsoft.com/windows/ie/default.asp)
- Netscape Communicator 6.2.3 (ou simplement Netscape 6.2.3) (version française) (http://www.netscape.com/)
- Mozilla 1.0 (version française) (http://www.mozilla.org/)
- Opera 6.04 (version française) (http://www.opera.com/)

Pour effectuer cette comparaison, nous allons considérer les 7 risques cités ci-dessus les uns après les autres, et pour chacun de ceux-ci, nous exposerons le comportement des 4 navigateurs mentionnés supra. Nous regarderons ce comportement sous 2 angles : le premier correspondra à la configuration par défaut du navigateur (l'utilisateur non averti surfera dans ces conditions, ce qui représente environ 95% des internautes) et les configurations qu'il est possible d'obtenir en modifiant les paramètres offerts à l'utilisateur dans les options de chaque browser.

II.1.1. Communication de l'adresse IP

Aucun des browsers considérés ne cachent l'adresse IP de l'utilisateur. Les options offertes à l'utilisateur ne lui permettent pas de remédier à cette situation. Indiquons tout de même que cela est assez normal puisque, comme nous le verrons, c'est davantage la tâche d'un proxy (à condition qu'il soit situé sur un autre ordinateur[52]) que d'un navigateur (cela lui étant totalement impossible dû à l'utilisation des protocoles TCP/IP sur lesquels la formulation de requêtes HTTP repose) ; ainsi cette situation n'est pas surprenante, bien qu'il faille tout de même en tenir compte.

II.1.2. Bavardage des browsers

Chacun des browsers étudiés est relativement bavard dans ses différentes configurations possibles (par défaut ou non)[53] : grâce aux langages de scripts (comme le JavaScript), un site Web peut assez facilement avoir accès à des informations telles que le type de système d'exploitation de l'utilisateur, le type et la langue du navigateur, les langues acceptées par l'utilisateur, la page référente, le temps local,...

Voici pour information, des captures très révélatrices du bavardage des navigateurs[54] :

- Opera 6.04 :

  User-Agent : Opera/6.04 (Windows XP; U) [fr]
  Accept : text/html, image/png, image/jpeg, image/gif, image/x-xbitmap, */*
  Accept-Language : fr, en
  Referer :
  http://webmail.fundp.ac.be/horde/imp/message.php3?index=1&array_index=0

[52] Pour une explication plus détaillée du « masquage » de l'adresse IP par un proxyserveur (qui, pour rappel, est une sorte d'intermédiaire entre un client et un serveur) et de son utilité, nous invitons le lecteur à consulter les deux prochaines sections.

[53] Néanmoins, certains le sont plus que d'autres, comme nous le verrons ci-après !

[54] Ces captures ont été obtenues grâce à un script situé à l'URL http://drt-144.droit.fundp.ac.be et mis au point par J.-M. Dinant. Nous n'avons indiqué ici que les variables d'environnement dignes d'intérêt pour notre propos (le script révélant l'entièreté du bavardage du navigateur y accédant).

- Mozilla 1.0 :
  User-Agent : Mozilla/5.0 (Windows; U; Windows NT 5.1; fr-FR; rv:1.0.0) Gecko/20020530
  Accept : text/xml, application/xml, application/xhtml+xml, text/html; q=0.9, text/plain; q=0.8,video/x-mng, image/png, image/jpeg, image/gif; q=0.2, text/css,*/*; q=0.1
  Accept-Language : en-us
  Referer :
  http://webmail.fundp.ac.be/horde/imp/message.php3?index=1&array_index=0

- MSIE 6.0 :
  User-Agent : Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
  Accept : image/gif, image/x-xbitmap, image/jpeg, image/pjpeg, application/vnd.ms-excel, application/vnd.ms-powerpoint, application/msword, */*
  Referer :
  http://webmail.fundp.ac.be/horde/imp/message.php3?index=1&array_index=0
  Accept-Language : fr-be

- Netscape 6.2.3 :
  User-Agent : Mozilla/5.0 (Windows; U; Windows NT 5.1; fr-FR; rv:0.9.4.1) Gecko/20020508 Netscape6/6.2.3
  Accept : text/xml, application/xml, application/xhtml+xml, text/html;q=0.9, image/png, image/jpeg, image/gif;q=0.2, text/plain;q=0.8, text/css, */*;q=0.1
  Accept-Language: fr-fr
  Referer :
  http://webmail.fundp.ac.be/horde/imp/message.php3?index=1&array_index=0

Le navigateur Opera offre dans ses options la possibilité de s'identifier comme étant un autre navigateur, dont MSIE 5.0 par exemple (qui correspond à la configuration par défaut) ; on obtiendrait alors :

User-Agent : Mozilla/4.0 (compatible; MSIE 5.0; Windows XP) Opera 6.04 [fr]

Ce qui, en fin de compte, en révèle tout autant !
Néanmoins, ce navigateur est le seul à disposer d'une option permettant de ne pas divulguer la page référente, mais elle est désactivée dans la configuration par défaut.

II.1.3. Hyperliens invisibles

Dans leur configuration par défaut, presque tous les browsers sont équivalents : ils affichent toutes les images[55] contenues dans les pages Web y compris celles provenant de sites tiers, sauf Mozilla qui n'affiche que les images du site d'origine et bloque les images des sites étrangers. Heureusement, ce n'est plus toujours le cas lorsque l'on configure les navigateurs grâce à plusieurs fonctionnalités qui nous sont offertes dans certains cas.

[55] Pour rappel, les hyperliens invisibles se « cachent » essentiellement derrière les images affichées à l'intérieur d'une page Web qui sont issues d'un ou plusieurs sites tiers, ces images sont la plupart du temps des bannières publicitaires ou des « images invisibles » dans le cas des Web bugs. Ainsi, ces images ne se situant pas sur le site que l'on visite mais sur un site extérieur, le navigateur ira les télécharger automatiquement sans en avertir l'utilisateur.

Voici les différentes options proposées par les browsers aux utilisateurs expérimentés :

- MSIE 6.0 : soit il affiche toutes les images sans distinction de leur source, soit il n'affiche aucune image[56]. Il n'y a aucune autre nuance.

- Opera 6.04 : il affiche soit toutes les images, soit uniquement celles préalablement contenues dans sa mémoire cache, soit aucune[5].

- Netscape 6.2.3 : Tout comme MSIE 6, il affiche soit toutes les images soit aucune[5].

- Mozilla 1.0 : Il peut afficher soit toutes les images, soit aucune, soit encore uniquement celles qui proviennent du site qui est en cours de visite (en excluant toutes celles provenant de sites tiers), cette dernière possibilité étant celle activée par défaut. De plus, l'utilisateur a la capacité d'être averti à chaque téléchargement d'image au cours du téléchargement d'une page Web et même de gérer une liste des sites dont on désire bloquer systématiquement les images[5].

Ainsi, nous avons constaté que tous les browsers ne sont pas égaux : Mozilla est ici certainement celui qui offre le plus d'options pour contrôler les effets des hyperliens invisibles, à condition tout de même de savoir le configurer comme il convient. Heureusement, en ce qui le concerne, la configuration par défaut correspond au meilleur choix (choix qu'il est le seul à proposer).

II.1.4. Redirection automatique

Dans toutes leurs configurations par défaut, tous les browsers sont équivalents : ils effectuent la redirection automatique sans que l'utilisateur en soit averti. Par contre, Opera est le seul à offrir une option permettant de la désactiver, mais la redirection automatique est activée par défaut. Opera se distingue donc des autres browsers sur ce point précis.

II.1.5. Cookies

En ce qui concerne les cookies[57] et leur gestion, chaque navigateur a un comportement différent à la fois dans sa configuration par défaut mais aussi dans les différentes configurations possibles qu'ils offrent à leurs utilisateurs.

---

[56] Remarque, pour information : Il y a moyen d'activer (configuration par défaut) ou de désactiver, si on le souhaite, les animations des images GIF, images qui constituent la plupart du temps des bannières publicitaires (mais pas toujours...). De telles animations étant lues en boucle, Mozilla et Netscape permettent également de les figer après une seule boucle.

[57] Pour information, signalons que les cookies nouvellement créés et en cours d'utilisation par le navigateur sont stockés systématiquement en mémoire vive ; les cookies persistants ne seront sauvegardés sur le disque dur que lorsque le browser sera quitté par son utilisateur. Par contre, les cookies de session, ayant un usage limité à la durée de la session établie entre le navigateur et les serveurs interlocuteurs qui les ont transmis, disparaîtront purement et simplement du système.

Configuration par défaut des browsers :

- MSIE 6.0 bloque les cookies tiers provenant de sites qui n'ont pas de stratégie de confidentialité compacte, de sites qui utilisent des informations d'identification personnelle sans le consentement de l'utilisateur et il restreint les cookies internes qui utilisent des informations d'identification personnelle sans le consentement de l'utilisateur[58]. En fait, avec Internet Explorer 6.0, Microsoft inaugure de nouvelles fonctions de confidentialité intégrées à son navigateur. Le logiciel reconnaît désormais les sites labellisés P3P[59] (Platform for Privacy Preferences), une sorte de certificat indiquant que le site respecte certaines règles élémentaires de protections de la vie privée. Le nouveau navigateur permet ainsi aux utilisateurs de n'accepter des cookies qu'en provenance de sites certifiés P3P[60].

- Opera 6.04 accepte tous les cookies.

- Netscape 6.2.3 accepte tous les cookies.

- Mozilla 1.0 accepte exclusivement les cookies du serveur visité (bloquant les cookies issus de sites tiers), et désactive l'emploi des cookies dans les e-mails et les forums de discussion.

Options avancées offertes à l'utilisateur averti :

- MSIE 6.0 : on peut soit utiliser la gestion automatique des cookies correspondant aux 6 configurations prédéterminées (on peut ainsi aller de l'acceptation de tous les cookies sans distinction jusqu'à tous les bloquer systématiquement, y compris ceux déjà présents sur le disque dur de l'utilisateur), soit ignorer cette gestion automatique des cookies en effectuant sa propre gestion personnalisée des cookies selon différents choix, ceux-ci sont : réception d'un avertissement et d'une demande d'acceptation (ou

---

[58] Plus concrètement, dans l'onglet confidentialité des options Internet du navigateur, le curseur est mis en position moyenne (position 3 sur 6, sachant que la position 1 correspond à accepter tous les cookies tandis que la position 6 correspond à refuser tous les cookies).

[59] Pour plus d'informations sur le standard P3P (et son rôle dans la protection des données à caractère personnel), une sous-section lui est consacré (cfr. Section II.4.4).

[60] Signalons au passage que Richard Smith, directeur technique de la Privacy Foundation, association américaine qui oeuvre pour le respect de la vie privée et des données personnelles, a révélé dernièrement une cohabitation malencontreuse de MSIE 6.0 avec le lecteur multimédia de Microsoft, Windows Media Player. Ce dernier présente une vulnérabilité qui permet à un hacker malintentionné de contourner les paramètres de sécurité d'Internet Explorer. Dans ce cas, un site Internet pourrait ainsi ne pas tenir compte des limitations P3P, alors que l'utilisateur se sent à l'abri. Microsoft a d'ailleurs fourni un patch pour corriger cette vulnérabilité, qui a été portée au grand jour en mars 2001. Le problème, c'est que peu d'utilisateurs penseront à appliquer ce patch au lecteur multimédia, fourni en standard avec Windows, pour corriger une faiblesse intervenant par ricochet au niveau du navigateur.
La vulnérabilité décrite par Smith réside dans le fait que Windows Media Player indique toujours le même numéro d'identifiant, par défaut, lorsqu'il se connecte à un site pour recevoir les flux de données audio/vidéo. Or, même si IE est capable de bloquer les cookies, ce numéro de série unique peut surtout servir à créer ce que Smith décrit comme un « super cookie », puisqu'en utilisant un code JavaScript sur une page HTML, un site peut saisir l'identifiant unique de Windows Media Player, ce numéro peut alors être utilisé tel un cookie par des sites Web pour pister la navigation d'un internaute.
Outre le patch de Microsoft, la solution est également de changer les paramètres par défaut de Windows Media Player. Il suffit ainsi de désactiver l'option « Autoriser les sites Internet à identifier le lecteur » dans les paramètres du lecteur (Outils/Options/Lecteur). Un numéro d'identifiant différent sera alors créé à chaque session.

de refus) à chaque réception d'un nouveau cookie (avec mémorisation du choix effectué, si on le désire), acceptation ou refus systématique des cookies internes (mêmes choix pour les cookies tiers) avec la possibilité de toujours accepter les cookies de session. De plus, il y a la possibilité de compléter les différentes gestions de cookies précédentes (qu'elles soient personnalisées ou non) au moyen d'une liste des domaines de sites Web désignant explicitement l'origine des cookies que l'on souhaite accepter ou refuser automatiquement.

- Opera 6.04 : on peut choisir de refuser tous les cookies, d'afficher les cookies reçus (ou seulement les cookies tiers), d'accepter les cookies uniquement des sites que l'on choisit, d'accepter tous les cookies sans distinction ou encore de n'accepter les cookies que du site que l'on est en train de visiter (et par voie de conséquence, de refuser les cookies de sites tiers). On peut également supprimer les nouveaux cookies en quittant le navigateur, demander l'affichage d'un avertissement pour les domaines et les chemins illégaux. En outre, Opera dispose d'un éditeur de filtres pour les cookies assez complet.

- Netscape 6.2.3 : on peut choisir de désactiver l'utilisation des cookies, de tous les accepter, de n'accepter que ceux issus du site Web visité ou encore d'être averti avant l'enregistrement d'un cookie sur le disque dur. On a aussi la possibilité de visionner l'ensemble des cookies qui sont stockés sur le disque dur (ce qui donne l'occasion de visualiser les informations qu'ils contiennent, dont leur date d'expiration et leur provenance, de les supprimer un par un ou tous à la fois, tout en mémorisant nos choix afin d'empêcher que les cookies que l'on a déjà supprimés soient à nouveau acceptés). Enfin, on peut sélectionner une fois pour toute les sites Internet qui peuvent/ne peuvent pas stocker des cookies sur l'ordinateur (tout en ayant la possibilité de gérer cette liste de sites dès qu'on le souhaite).

- Mozilla 1.0 : il reprend toutes les options offertes par Netscape 6.2 et en rajoute quelques autres. Ainsi, on peut fixer soi-même la durée de vie maximale de tous les cookies que l'on reçoit (soit uniquement pour la durée de la session en cours, soit pour une durée que l'on définit en jours) et désactiver l'utilisation des cookies dans les e-mails et les forums.

On constate donc que Mozilla 1.0 est le browser qui offre la plus grande flexibilité au travers de ses options de configuration tout en étant le plus « sûr » dans sa configuration par défaut.

II.1.6. Scripting côté client

La configuration par défaut de tous les browsers testés active le scripting côté client. Heureusement, les utilisateurs peuvent modifier plusieurs paramètres de configuration pour l'adapter selon leurs envies. De nouveau, nous verrons que dans ce domaine, tous les navigateurs présentés ne sont pas égaux.

Voici la configuration par défaut des différents navigateurs :

- MSIE 6.0 : si le moteur virtuel Java de Microsoft ou le plug-in Java 2 de SunMicrosystems ne sont pas installés sur la plateforme Windows XP (qui n'incorpore ni l'un ni l'autre d'origine), les applets Java ne peuvent pas être exécutées. Par contre, une fois un des deux interpréteurs installés, par défaut l'exécution des applets Java sera activée, mais selon l'option « Haute sécurité ».

En ce qui concerne les autres types de scripting, ils ne nécessitent pas d'installation préalable et sont activés par défaut. Le téléchargement de contrôles ActiveX signés se fait sur demande (et est interdit s'ils ne sont pas signés) ; de même, l'exécution des contrôles ActiveX et des plug-ins a lieu à la demande de l'utilisateur. Par contre, les contrôles ActiveX reconnus sûrs pour l'écriture de script sont activés. Ces options correspondent au niveau de sécurité moyen du navigateur, niveau sensé fournir une navigation sûre mais néanmoins pratique qui est adaptée à la plupart des sites Internet. D'un point de vue multimédia, les sons et les vidéos sont lus dans les pages Web.

- Opera 6.04 active la lecture des sons et des vidéos, l'utilisation du JavaScript ainsi que des plug-ins. En fait, il existe deux versions de ce navigateur, une avec Java l'autre sans. Bien sûr, la version sans Java est incapable d'exécuter des applets Java. Par contre, dans la version avec Java, l'exécution des applets est activée par défaut.

- Netscape 6.2.3 active Java et JavaScript dans Navigator, mais aussi le JavaScript dans les e-mails et les forums.

- Mozilla 1.0 autorise l'exécution d'applets Java. Le JavaScript est activé dans le Navigator mais pas dans les e-mails et les forums. Par contre, il interdit aux codes écrits en JavaScript d'ouvrir des fenêtres non désirées et de lire, modifier ou créer des cookies.

Quelles sont les possibilités de configuration accessibles aux internautes informés ?

- MSIE 6.0 permet d'activer/désactiver l'emploi du plug-in Java de Sun ou du moteur virtuel Microsoft. De même, on peut autoriser/stopper la lecture des sons et des vidéos dans les pages HTML. Enfin, on peut personnaliser le niveau de sécurité lors de la navigation par un choix entre 4 niveaux prédéfinis (de faible, en passant par moyennement bas et moyen, jusqu'à élevé), chacun de ces niveaux correspondant à des réglages particuliers (ainsi dans le niveau bas, la protection est minimale, les contrôles ActiveX et les contenus actifs sont exécutés sans en avertir l'utilisateur, ce niveau est donc adapté aux sites dont on a une confiance absolue ; par contre, dans le niveau élevé, la navigation est la plus sûre mais la moins pratique, toutes les fonctionnalités représentant le plus de risques sont désactivées, ce niveau est adapté aux sites dont le contenu est offensif). Mais on peut aussi établir son propre niveau de sécurité en paramétrant les nombreuses options disponibles (signalons qu'elles sont tout de même réservées aux internautes qui savent évaluer l'impact de leurs décisions en terme de risques potentiels vis-à-vis desquels ils se protègent ou non).

- Opera 6.04 : on peut activer/désactiver l'usage du JavaScript, d'applets Java, ou des plug-ins, mais aussi des sons et des vidéos, par les sites Web.

- Netscape 6.2.3 : de même, on peut autoriser/interdire les applets Java, ou le code JavaScript (pour le Navigator et/ou dans les e-mails et les forums).

- Mozilla 1.0 : il possède tous les réglages présents dans Netscape mais permet également un contrôle plus précis des actions autorisées/interdites par le JavaScript incorporé dans les pages Web.

Pour le scripting côté client, aucun des navigateurs ne se démarque réellement : ils ont tous une configuration par défaut similaire et possèdent des réglages plus ou moins équivalents. Toutefois, MSIE 6 propose davantage de possibilités de personnalisation vis-à-vis du scripting que les autres browsers ; cependant, il faut faire attention aux conséquences lors de la modification de ces paramètres car ils influent fortement sur la sécurité.

II.1.7. Mémoire cache

Remarque préliminaire : En ce qui concerne la mémoire cache (ainsi que les cookies stockés sur le disque dur), il y a des différences notables qui proviennent du système d'exploitation utilisé par l'internaute. En effet, sur un système d'exploitation non multi-utilisateur (Microsoft Windows 9x, par exemple), si plusieurs personnes (une famille, des collègues de bureau,...) se connectent à Internet, elles auront une mémoire cache commune (même répertoire de stockage des fichiers temporaires, mêmes fichiers de cookies,...) et donc partagée. Cela implique que les cookies présents sur le système ne permettront pas de profiler une personne de façon spécifique, mais l'ensemble des utilisateurs. Les sites qui utilisent des cookies n'auront donc accès qu'à un profil collectif, une sorte de moyenne des habitudes des utilisateurs. Ces cookies seront par conséquent moins révélateurs du comportement d'un utilisateur en particulier, et moins pertinents pour le site qui l'utilise.
Par contre, sur un système multi-utilisateur (Microsoft Windows NT/2000/XP ou Linux, pour ne citer qu'eux), si plusieurs utilisateurs se connectent au système, l'administrateur de celui-ci leur a attribué à chacun un compte utilisateur différent (identifié par un login et pouvant être protégé par un mot de passe). Dans ce cas, chaque utilisateur possède sa propre mémoire cache (son répertoire de stockage des fichiers temporaires, ses propres fichiers de cookies,...) liée à son compte utilisateur particulier. Comme il n'y a plus de partage de la mémoire cache, la remarque précédente concernant les systèmes non multi-utilisateur n'est plus d'application et les cookies retrouvent leur aspect révélateur du comportement d'un utilisateur spécifique, sauf bien sûr si plusieurs personnes utilisent le même compte utilisateur (mais elles perdent par la même occasion les avantages propres à l'utilisation de comptes utilisateurs différents, comme la confidentialité des documents,...).

Tous les navigateurs sont relativement équivalents lorsqu'il s'agit de mémoire cache. En effet, ils offrent tous cette fonctionnalité, et cette dernière est toujours activée dans la configuration par défaut, seule change la taille allouée à cette mémoire. Nous ne parlerons pas ici des options permettant de modifier le fonctionnement proprement dit du cache, en particulier son mode de gestion.

Configuration par défaut :

- MSIE 6.0 : l'antémémoire est activée, sa taille est relativement grande (cela dépend de la taille du disque dur). Le navigateur utilise un répertoire par défaut pour stocker tous les fichiers et nous en indique son emplacement. L'historique des liens visités est conservé pendant 20 jours.

- Opera 6.04 : tout comme MSIE, le cache est activé, seulement sa taille est beaucoup plus petite (2 Mo). Par conséquent, les fichiers sont plus souvent effacés pour être remplacés par de plus récents, et donc le contenu de la mémoire cache sera moins révélateur de l'entièreté de la navigation lors d'une session sur Internet. Signalons que par défaut, Opera enregistre dans son antémémoire tous les types de fichiers (documents, images, ...). La gestion des historiques est plus fine, avec la distinction entre les liens visités (maximum 500 adresses), les adresses entrées au moyen de la

barre d'adresses (maximum 200 adresses) et les adresses des pages obtenues par les actions Précédent et Suivant (maximum 50 adresses). Ces trois types d'historique sont activés.

- Netscape 6.2.3 : le cache est également activé, et sa taille par défaut est de 50 Mo. L'historique des liens visités n'est conservé que pendant la journée en cours (ce qui correspond à 0 jour). L'historique de session (càd la liste des pages consultées précédemment pour une fenêtre de navigateur ouverte, accessible à l'aide des boutons Précédent et Suivant) a une capacité de 50 pages.

- Mozilla 1.0 : Même configuration que Netscape. Néanmoins, il indique l'emplacement du répertoire servant d'antémémoire.

Configurations avancées :

- MSIE 6.0 : Un bouton permet d'effacer l'entièreté des fichiers contenus dans l'antémémoire (avec ou sans suppression du contenu hors connexion), et un autre permet de supprimer les cookies stockés. On peut modifier la taille maximale de la mémoire cache ainsi que son emplacement. En outre, on peut en afficher le contenu. Un bouton permet d'effacer à la fois le contenu de l'historique des liens visités et celui des adresses entrées grâce à la barre d'adresses. Enfin, une option permet de régler le nombre de jours de conservation des liens dans l'historique (0 jour correspondant à la durée de la journée entamée), et une autre permet de vider le cache en quittant le navigateur.

- Opera 6.04 : on peut modifier la taille du cache, activer/désactiver les différents types de contenu (documents, images ou autre) stockés. Un bouton permet d'effacer tout le contenu de l'antémémoire et une option active la suppression automatique des fichiers enregistrés lorsque l'on quitte le navigateur. Il y a moyen également de paramétrer les trois différents types d'historique (les activer ou désactiver au choix, modifier le nombre maximal d'adresses enregistrables). Un bouton permet d'effacer le contenu de l'historique des liens visités, un autre celui des adresses entrées.

- Netscape 6.2.3 : Un bouton permet de vider le contenu de la cache et une option permet d'indiquer sa taille maximale sur le disque. On peut spécifier le nombre de jours pendant lesquels les liens visités seront conservés dans l'historique correspondant. Un bouton permet de supprimer le contenu de ce dernier, un autre a pour but d'effacer des adresses entrées. On peut également modifier le nombre de liens dans l'historique de session.

- Mozilla 1.0 : Mêmes options que Netscape. La seule différence, c'est que l'on peut changer l'emplacement du répertoire où sont stockés les fichiers de la mémoire cache.

II.1.8. Quelques remarques diverses

Il est intéressant de constater qu'Opera dispose d'une fonction, que l'on ne retrouve pas dans les autres navigateurs testés, permettant d'effacer, selon les desiderata de l'utilisateur, les multiples informations que le navigateur a stockées sur le disque dur au cours de la session passée sur Internet[61]. Pour les autres navigateurs, il faut se rendre dans les différents menus de configuration et trouver les différents boutons permettant de supprimer les fichiers présents dans la mémoire cache, l'historique des liens visités, les URLs entrées dans la barre d'adresses, les cookies... Il est dommage qu'ils ne soient pas tous rassemblés dans le même onglet, ou dans des onglets voisins. Opera est également le seul à posséder une option (malheureusement désactivée par défaut) pour empêcher les fenêtres pop-up d'apparaître, ou pour les ouvrir en arrière plan. Enfin, il dispose d'un menu déroulant intitulé « Préférences rapides » qui permet de configurer en quelques clics de souris les options importantes concernant la sécurité et la protection de la vie privée[62] ; c'est l'outil idéal pour les novices qui craignent de s'égarer dans les nombreux menus de configuration et qui veulent aller droit à l'essentiel.

Pour Netscape et Mozilla, l'utilisateur peut s'informer et être sensibilisé à certains aspects concernant la protection de sa vie privée sur Internet en lisant un petit texte[63] expliquant le degré de confidentialité auquel il est en droit de s'attendre lorsqu'il surfe sur le Web et comment il peut contrôler les informations transmises à son sujet. Ainsi, ce texte décrit en bref les principales techniques au moyen desquelles un site peut collecter des informations sur un internaute particulier.

II.1.9. Conclusion

Pour conclure brièvement cette analyse, nous estimons que Mozilla 1.0 est le navigateur le plus respectueux de la vie privée de ses utilisateurs (surtout en configuration par défaut, càd celle utilisée par un utilisateur novice). Le « navigateur idéal » serait un browser comme Mozilla avec certaines options supplémentaires provenant d'Opera comme la possibilité de refuser les pop-ups, d'effacer en un seul geste toutes les informations enregistrées par le navigateur pendant son utilisation, la capacité de désactiver la redirection automatique ainsi que la transmission de l'adresse de la page référente et l'incorporation d'un menu de préférences rapides.

Pour terminer, remarquons que la version testée d'Opera, la 6.04, est plus nuisible, en configuration par défaut, pour la vie privée de ses utilisateurs que la version 5.12 du même navigateur. Par contre, la version testée de Mozilla, la 1.0, est plus respectueuse, en configuration par défaut, de la vie privée que la version précédente 0.9.6. Donc, les conclusions que l'on peut tirer à partir de cette étude sont acceptables a priori strictement pour

---

[61] Il s'agit d'effacer, selon les choix de l'utilisateur, tous les cookies, la totalité du contenu de la mémoire cache (avec ou sans les documents cryptés), les mots de passe mémorisés et/ou l'historique des liens visités, des URLs entrées ou encore des fichiers transférés. Signalons que l'utilisation de cette fonctionnalité conduit à fermer le navigateur.

[62] Ce menu permet d'activer/désactiver les fenêtres pop-up, les animations des images GIF, les sons et les vidéos dans les pages Web, l'utilisation du JavaScript, de Java ou des différents plug-ins, l'emploi des cookies, la transmission de l'adresse du site d'où l'on vient, ou encore de modifier la signature d'identification du navigateur.

[63] Il pourra y accéder dans les options de configuration (onglet confidentialité et sécurité - cookies).

les versions des navigateurs qu'elle concerne et ne doivent pas, d'après nous, être trop extrapolées, puisque les configurations par défaut des navigateurs évoluent avec les versions successives (en bien ou en mal vis-à-vis de la protection de la vie privée).

Une façon surprenante pour tout connaître de la configuration de son navigateur est de consulter l'outil en ligne ***Browser Spy*** (http://gemal.dk/browserspy/), qui permet de capter une foule d'informations révélées par le navigateur à l'insu de son utilisateur (pour ce faire, il utilise JavaScript ; pour le rendre muet, il faut donc désactiver cette fonctionnalité dans son browser sinon c'est une bonne raison pour se rendre compte de la nécessité d'une telle opération pour la préservation de sa vie privée !).

II.2. Les limites de certaines techniques de collecte de données personnelles

II.2.1. Les fichiers d'audit

La meilleure manière, pour un internaute, de mettre en échec le profilage sur base des fichiers d'audit établis à partir des adresses IP est que cet internaute « cache » sa propre adresse IP ! En toute logique, cela signifie que l'adresse IP de l'internaute, que le site Web « pisteur » reçoit lors de ses différentes requêtes, n'est pas la sienne ou pas seulement la sienne.

Ceci peut se produire, par exemple, dans les cas suivants :

a) Le cybercafé

Supposons qu'un internaute se connecte en utilisant un ordinateur en libre service, comme dans un cybercafé, ou à partir d'une borne située dans un lieu public ; un autre internaute aura (ira) peut-être visité le même site avant lui (après lui) à partir du même ordinateur et aura donc laissé (laissera) la même trace que lui ; ces traces seront alors assimilées comme provenant d'une seule personne (le profilage sera donc faussé). Régulièrement les fichiers d'audit seront remis à zéro, par conséquent, si plusieurs personnes se sont connectées avant cela, toutes leurs connexions (et leurs traces) sont amalgamées.

b) Le firewall ou le NAT

Si un internaute se connecte à travers un réseau local, il est possible qu'un firewall[64] soit en place et cache sa véritable adresse IP en la remplaçant par une autre. Plus précisément, le firewall situé à la sortie du réseau local[65], de son entreprise par exemple, substitue une autre

[64] Il ne s'agit pas ici d'un firewall software installé sur la machine de l'internaute, mais d'un firewall hardware situé sur le lien reliant le réseau local à Internet.

[65] Ce type de configuration est relativement courant dans les grandes structures (universités, grandes entreprises, etc.). On la trouve le plus souvent dès lors qu'un réseau local, plutôt qu'un PC isolé, est connecté à Internet. En effet, de petits et de grands ordinateurs commerciaux à usage spécifique sont des firewalls destinés uniquement à protéger l'intégrité du réseau local de leur organisation. Les firewalls fournissent cette protection en constituant une porte virtuelle vers le réseau d'ordinateurs interne de la compagnie. Alors que les employés peuvent passer à travers cette porte pour accéder à Internet, les utilisateurs externes sont bloqués et ne peuvent avoir accès directement aux ordinateurs de la société. Ces firewalls empêchent les hackers de frauder le site Web de la compagnie ou d'atteindre des fichiers internes sensibles.

adresse IP à la sienne et se charge lui-même de la distribution des paquets à leurs bons destinataires sur le réseau local ou de leur envoi sécurisé sur Internet. Il peut ainsi utiliser une même adresse IP pour plusieurs personnes. La trace laissée dans le fichier d'audit d'un serveur sur Internet est donc celle du firewall et non celle de l'internaute. Un tel dispositif est généralement installé pour des raisons de sécurité, afin d'éviter toute intrusion extérieure sur le réseau local et de cacher les adresses IP réelles des ordinateurs du réseau (ou de les traduire si les adresses IP employées au sein du réseau local sont invalides sur Internet). Un autre intermédiaire pouvant relier un réseau local à Internet est ce qu'on appelle un NAT (Network Adress Translator) dont l'unique tâche est d'effectuer une traduction d'adresses IP[66], sans effectuer de filtrage sur les paquets IP transitant à travers lui (il ne constitue donc pas un moyen de protection vis-à-vis des attaques extérieures.

Néanmoins, lorsqu'un tel dispositif est mis en place, tout ce qu'un serveur Web peut faire en terme de traçage d'adresse IP et de fichiers d'audit, un firewall (ou un NAT ou même n'importe quel autre intermédiaire) peut le faire aussi ! Ainsi, ce n'est plus le serveur Web qui peut suivre un internaute à la trace mais l'administrateur du firewall, probablement situé dans son entreprise, qui peut contrôler ses allées et venues sur le réseau. Et dans ce cas-là, non seulement il disposera de son parcours sur un serveur particulier, mais il disposera de la même manière de toutes ses requêtes et par conséquent, sera en mesure de reconstituer entièrement les multiples sessions qu'il a effectuées puisque toutes ses requêtes auront forcément transité par le firewall, point d'accès unique du réseau local à Internet.

*c) **Le proxyserveur***

Un serveur proxy peut être situé à n'importe quel endroit du réseau mais l'est généralement chez un fournisseur d'accès. Lorsqu'un internaute exécute une requête vers un serveur du réseau, le proxy en question vérifie si la même requête a déjà été faite précédemment. Si ce n'est pas le cas, il la renvoie sur le réseau, et transmet son résultat à l'internaute en question (la page HTML demandée par exemple) après l'avoir enregistré sur son propre disque dur afin de satisfaire de futures requêtes. Ainsi, si l'internaute ou un autre client du même fournisseur d'accès lance la même requête peu de temps après, c'est le serveur proxy qui y répondra en allant chercher son résultat à l'endroit où il l'a stocké. D'où un gain de temps et une économie de bande passante pour le fournisseur d'accès. C'est exactement le même principe que la mémoire cache que l'on trouve dans les browsers Web mais cette fois accessible à un plus grand nombre de personnes. Chaque page, image ou autre objet du Web importé d'Internet par un membre de l'organisation à laquelle appartient le proxyserveur est stockée sur son disque dur, qui fait office d'antémémoire, et sera immédiatement disponible pour un autre membre de cette organisation.

Cette situation implique deux conséquences. Tout d'abord, c'est la trace du proxy qui figurera dans le fichier d'audit du serveur Web visité (puisque le serveur proxy ne transmettra, normalement[67], pas l'adresse IP de l'internaute au site Web). De plus, il peut être capable de

---

[66] Cette traduction permet à plusieurs personnes de partager un unique accès à Internet désigné par une seule adresse IP « valide », celles des ordinateurs des personnes partageant l'accès étant des adresses de réseau local considérées comme « invalides » sur Internet.

[67] Normalement, car certains proxies sont plus bavards qu'on ne le croit... Il faut donc s'assurer que le proxy utilisé est « non transparent », càd qu'il ne transmet pas les adresses IP de ses utilisateurs aux serveurs sur lesquels ceux-là souhaitent se connecter ! Signalons que cette notion de « transparence » d'un proxy est inscrite dans la norme HTTP elle-même (RFC 2616 : « *A "transparent proxy" is a proxy that does not modify the request or response beyond what is required for proxy authentication and identification. A "non-transparent proxy" is a proxy that modifies the request or response in order to provide some added service to the user agent, such as*

filtrer le bavardage du navigateur et, du fait qu'un serveur proxy gère le protocole HTTP, les cookies installés dans l'en-tête des requêtes peuvent aussi être effacés, modifiés ou stockés par lui[68]. Ensuite, le proxy pourra très bien conserver l'intégralité de l'historique de toutes les requêtes de l'internaute. Si le proxy est situé dans son entreprise, la configuration est la même que celle envisagée ci-dessus dans le cas du firewall. S'il est chez un fournisseur d'accès, c'est ce dernier qui dispose de l'ensemble des informations concernant toutes les transactions de ses clients.

d) Des sites Web anonymiseurs

Dernier cas de figure : certains sites Web proposent de jouer le rôle d' « anonymiseur »[69]. Les sites Web d'anonymat[70], comme http://www.anonymizer.com[71] par exemple, utilisent des serveurs proxies pour masquer la navigation d'un internaute. Concrètement, un internaute se connecte à un site Web anonymiseur A. Ce dernier lui propose de taper l'URL d'un autre site B dans une boîte de saisie textuelle d'une page Web de A. Puis

---

group annotation services, media type transformation, protocol reduction, or anonymity filtering », cfr. [23]) et est donc tout à fait autorisée (et couramment utilisée).

[68] Toutefois, le filtrage par l'intermédiaire de proxy est aujourd'hui mis à mal par la présence du langage JavaScript qui permet l'utilisation de cookies dans le corps du langage HTML et non plus dans l'entête HTTP.

[69] On entend souvent par là le fait de cacher l'adresse IP réelle de l'internaute utilisant le système d'anonymisation. Néanmoins, il peut aussi s'agir d'une protection des données à caractère personnel plus étendue (filtrage des cookies, chiffrement de la communication,...) si le système en question (un logiciel spécialisé installé sur la machine de l'internaute, par exemple) le permet. Pour plus de clarté (et par abus de langage), j'entendrai dorénavant par « être anonyme » sur Internet le fait de ne pas communiquer son adresse IP réelle aux serveurs des sites que l'on souhaite consulter (on conviendra, suite à la discussion du premier chapitre, que cet « anonymat » ne permet pas suffisamment de protéger sa vie privée). La raison d'être de cette expression est que l'adresse IP est l'identité unique d'un utilisateur sur Internet pendant le temps de sa connexion au réseau (son nom, son adresse postale n'ayant aucune valeur identifiante sur le réseau). Il est important aussi de remarquer que l'on n'est jamais anonyme pour l'ensemble du réseau sur lequel transitent nos requêtes. Il faut donc savoir par rapport à quel serveur on est anonyme (par exemple, il n'y a pas d'anonymat entre l'ordinateur de l'internaute et le site Web anonymiseur utilisé, par contre l'internaute est anonyme vis-à-vis des sites qu'ils consultent par l'intermédiaire du site anonymiseur car il ne leur communique pas sa propre adresse IP, celle-ci étant cachée par le site anonymiseur ; cela signifie aussi que si les sites visités ne peuvent pas pister l'internaute sur base de son adresse IP, le site anonymiseur lui le peut sans aucune difficulté !).

[70] http://anonymizer.secuser.com, http://nonymouse.com/, http://www.anon.de/ et http://www.rewebber.de/ (sans SSL, « https:// » pour connexion SSL sécurisée), https://proxy1.autistici.org/ (avec SSL), http://www.surfola.com/, http://www.megaproxy.com/_secure/ (sans SSL, « https:// » pour connexion SSL sécurisée), https://proxy.magusnet.com/ (avec SSL), http://www.guardster.com/ ou encore http://anonymouse.is-4u.de

[71] Pour Anonymizer, il suffit juste de taper l'URL du site que l'on aimerait visiter anonymement dans la boîte textuelle intitulée « Free Anonymous Surfing » au sommet de la page et de sélectionner le bouton « Go ». On voit alors une page nous demandant d'accéder à notre compte personnel. Si on est un utilisateur Premium (utilisateur payant), on entre ses informations d'accès au compte et on est immédiatement envoyé sur le site que l'on a demandé. Si on désire utiliser le service gratuitement, on doit juste sélectionner le bouton « Surf for free ». En cliquant sur ce bouton, on obtient la page demandée après un petit délai, nécessaire pour que Anonymizer donne la priorité de son trafic aux utilisateurs Premium.
Une fois qu'on est arrivé sur le site que l'on a demandé, Anonymizer place un panneau de contrôle au sommet de la page visitée. Ce panneau nous permet d'accéder à d'autres pages anonymement, sans avoir besoin de retourner sur le site Web de Anonymizer. En outre, n'importe quel lien hypertexte sur lequel on clique pendant notre navigation en utilisant Anonymizer est aussi traité de façon anonyme.

le serveur A envoie la requête vers B, lequel transmet à A le résultat de sa requête. A re-dirige alors ce résultat vers l'internaute.

Le fichier d'audit de B contiendra donc l'adresse IP de A, et seul le fichier d'audit de A contiendra l'adresse IP de l'internaute. Pour l'identifier, l'administrateur du serveur B devrait demander au serveur A un extrait de son fichier d'audit, ce que A, a priori, refuserait de faire (c'est du moins ce que ce type de serveurs annoncent). Il faut noter tout de même que si l'anonymiseur a tracé l'internaute, il est en mesure de savoir à quoi il s'intéresse et, mieux encore, il sait qu'il a souhaité rester anonyme.

Il existe également certains sites Web anonymiseurs qui offrent certains services supplémentaires et ne se limitent pas à cacher l'adresse IP d'un internaute. Comme nous l'avons évoqué ci-dessus, un serveur proxy est capable de modifier le bavardage des navigateurs ou de contrôler les cookies. Ainsi, le site ***the-cloak*** (http://www.the-cloak.com/anonymous-surfing-home.html) supporte les protocoles FTP, HTTP et HTTPS[72] au moyen de son proxy d'anonymisation (qui masquera donc l'adresse IP de l'internaute l'utilisant). Ce dernier peut être configuré par l'internaute pour filtrer[73] les cookies (ou les supprimer définitivement) et même bloquer sélectivement le JavaScript[74] ou les applets Java (l'internaute ayant le choix de supprimer ou non ce type de contenu ainsi que certains types d'objets insérés dans les pages Web). Depuis peu, le site permet également de bloquer les bannières publicitaires (au moyen d'une liste répertoriant les serveurs de publicité). Le service fournit par ce site est gratuit.
Un autre exemple est le site ***subDimension*** (http://anonymizit.subdimension.com/cgi-bin/subnet/nph-subnet-demo.cgi) qui permet également de surfer anonymement sur Internet (support des protocoles HTTP et FTP mais pas HTTPS) mais également de filtrer les cookies (et même de les gérer en ligne), les bannières publicitaires, les scripts et de cacher le champ transmis par le navigateur indiquant la page référente. Malheureusement, le service offert par ce site est payant et la version démo est limitée au site de Google. Enfin, le site ***Ponoi*** (http://www.ponoi.com/) offre un service quasiment identique, accessible n'importe où et n'importe quand. Il permet de surfer anonymement en utilisant une connexion sécurisée, de gérer en ligne ses mots de passe et de stocker des fichiers sur leur serveur protégé. Selon leurs dires, lorsque l'on utilise leur service, personne, pas même eux, ne peut savoir qui nous sommes ou faire le lien entre notre identité et ce que nous faisons sur le Web. Ce service est payant mais la navigation anonyme peut être essayée gratuitement.

On en conclut donc que la seule limite des fichiers d'audit, c'est lorsque les administrateurs des serveurs par lesquels transitent les requêtes d'un internaute ont décidé de ne pas exploiter cette fonctionnalité ! Si ce n'est pas le cas, il est impossible d'y échapper totalement car le fonctionnement du protocole TCP/IP rend indispensable l'utilisation d'adresse IP lors de l'échange d'informations entre un internaute et un serveur. Cependant, il est possible de faire en sorte que les informations collectées par le serveur vis-à-vis duquel un

[72] Il est déconseillé d'utiliser le HTTPS pour effectuer du e-commerce à partir de ce site ou transmettre des informations sensibles à un site sécurisé (mot de passe, numéro de carte de crédit,...) puisque ce dernier peut accéder à la totalité des données échangées.

[73] Le filtrage des cookies par ce site consiste en la modification des cookies persistants en cookies de session. D'autre part, les cookies ne sont pas transmis à l'internaute et ils restent uniquement dans la mémoire du proxy, jusqu'à la fin de la session. Cette option est utile pour surfer sur les sites qui exigent l'emploi de cookies.

[74] Une option permet de filtrer le JavaScript pour le rendre « sain », mais il n'y a aucune garantie quant à la complète fiabilité de l'opération.

internaute souhaite être anonyme ne correspondent pas à l'ordinateur utilisé par cet internaute en plaçant un intermédiaire substituant alors sa propre adresse IP à celle de son ordinateur. Mais cette solution ne fait que remonter d'un cran la conséquence : ce que ne peut plus faire l'administrateur du serveur auquel l'internaute accède, l'administrateur du système intermédiaire, lui, le peut.

Il ne reste donc qu'un dernier cas dans lequel il est beaucoup plus difficile d'être tracé, c'est lorsque l'internaute effectue ses connexions grâce à un cybercafé ou à partir d'une borne publique d'accès à Internet. Là, même si tout le dispositif d'audit est en place, comment savoir qui a utilisé cet ordinateur à ce moment là ? Il faudrait avoir placé l'établissement sous surveillance vidéo...

II.2.2. Les cookies

En bref, voici les avantages et inconvénients des diverses solutions que l'on retrouve le plus fréquemment sur les navigateurs récents pour limiter l'usage des cookies :

1) Radicalement

Pour être absolument sûr de ne pas traîner de cookies indésirables, la méthode la plus radicale consiste à utiliser des navigateurs incapables de les traiter. C'est le cas des plus anciens et de certains autres qui ont volontairement choisi de ne pas exploiter cette fonctionnalité. Mais ce choix conduit également à ne plus pouvoir utiliser les évolutions de ces navigateurs, ni leurs avantages sur l'interactivité offerte à son utilisateur par les sites Web les plus sophistiqués.

2) A la carte

Les navigateurs de dernière génération proposent dans leurs options de paramétrage de contrôler les cookies en direct, au moment où ils arrivent. Ce faisant, l'internaute peut autoriser les sites à utiliser des cookies seulement après en avoir été averti par le navigateur. Un message s'affichera alors à chaque tentative de placement d'un cookie par un site. Ce message identifie le site (qui n'est pas nécessairement celui que l'internaute est en train de visiter, cas des cookies étrangers) et lui demande s'il autorise le placement du cookie. Le navigateur lui demande également s'il souhaite mémoriser sa décision pour la provenance de ce cookie. S'il accepte le cookie d'un site et spécifie la mémorisation de ce choix, le navigateur autorisera désormais automatiquement le stockage des cookies en provenance de ce site sans l'en avertir. Inversement, si l'internaute refuse le cookie et demande également de mémoriser cette décision, le navigateur refusera à l'avenir toutes les tentatives de stockage des futurs cookies pour ce site. Concrètement, lorsqu'un cookie est envoyé par le serveur visité par l'internaute au navigateur, une fenêtre s'ouvre et indique un message du type « Le serveur *www.machinchose.com* vous a envoyé un cookie dont le contenu est "*xxx yyy zzz*". Souhaitez-vous l'autoriser ? ». Si l'internaute répond par la négative, le cookie n'est pas enregistré, et le navigateur ne le prend pas en compte. Cette solution présente un avantage et un inconvénient : d'un côté, l'on prend enfin en compte l'avis de l'internaute (celui-ci pouvant voir explicitement le contenu du cookie qui lui est envoyé, ainsi que sa provenance et sa durée de validité, pour baser sa décision), de l'autre, les nombreuses annonces de réception de cookies parasitent en permanence sa navigation sur Internet à tel point qu'il finit par passer plus de temps à accepter ou refuser des cookies qu'à consulter le site. Par conséquent, il s'en lasse très vite et finit par préférer l'acception (ou le refus) automatique de tous les cookies.

3) Radicalement bis

C'est pourquoi il est également possible dans les derniers navigateurs de refuser purement et simplement tous les cookies. Mais il faut tout de même remarquer que le fait d'interdire systématiquement l'emploi des cookies peut entraîner des désagréments lors de la navigation sur certains sites Web (et ils sont de plus en plus nombreux), puisque l'internaute se prive ainsi des cookies dont la tâche est d'améliorer son confort lors de ses différentes visites, ou même empêcher purement et simplement l'accès aux sites dont le fonctionnement repose irrémédiablement sur eux.

4) Manuellement

Néanmoins, il peut être utile de trouver un juste milieu en autorisant certains sites à placer des cookies (les sites de webmail, par exemple), l'interdire à d'autres (connus pour s'adonner à des pratiques de marketing douteuses) et, par exemple, être invité à donner notre permission pour tous les autres sites. Ainsi, les navigateurs actuels proposent désormais la possibilité d'indiquer le nom de domaine des sites dont on souhaite accepter ou refuser une fois pour toute les cookies. Cette solution peut être pratique pour un internaute lorsque, après avoir utilisé pendant un certain temps la fonction « à la carte » du browser lors de la navigation sur les sites qu'il visite le plus souvent, il a pu déterminer les sites tiers à l'origine des bannières publicitaires, les sites exigeant l'utilisation de cookies pour leur fonctionnement, etc. Après cette période d'observation, l'internaute s'est fait une bonne idée des sites dont il est prêt à accepter les cookies et de ceux dont il ne souhaite plus recevoir de cookies parasites. Ainsi, cette fonction pourra être utilisée par l'internaute, ce qui lui permettra de ne plus recourir à la fonction « à la carte » (avec ses inconvénients) qui sera ainsi remplacée efficacement sans devoir effectuer un choix catégorique (tout accepter ou tout refuser). La difficulté est qu'il est impossible pour un internaute de répertorier tous les sites qu'il est susceptible de visiter un jour ou l'autre ; ainsi, cette solution est surtout valable pour les sites les plus souvent consultés par cet internaute. Pour les autres, le problème reste posé !

5) Manuellement bis

Une dernière option consiste à accepter tous les cookies, et à nettoyer périodiquement le(s) fichier(s) qui les contiennent. Pour ce faire, on peut ouvrir ce(s) fichier(s) avec un éditeur de texte standard, de le(s) modifier (selon les préférences de l'internaute) et de le(les) enregistrer en read-only pour empêcher toute modification ultérieure, en ayant au préalable contrôlé ou effacé les informations qu'il(s) contient(contiennent). Mais on peut également supprimer purement et simplement ce(s) fichier(s). Cette opération devra être répétée régulièrement car de nouveaux fichiers (correspondant à de nouveaux cookies) font souvent leur apparition à chaque nouvelle session sur Internet.

Cette démarche est particulièrement constructive car elle permet à chacun de voir très clairement quels serveurs utilisent des cookies, en particulier ceux dont les cookies ont un contenu incompréhensible, ceux qui se servent de cookies dans un but de convivialité par exemple (à condition que le contenu du cookie soit intelligible), etc. Mais il est vrai qu'à la longue, cette exploration peut être lassante ; c'est pour cette raison que des logiciels dévoués uniquement à la gestion des cookies ont fait leur apparition (nous en reparlerons dans la section « Autres technologies de renforcement de la protection de la vie privée »).

Le gestionnaire de cookies intégré à la plupart des navigateurs permet à l'internaute d'afficher la liste des cookies enregistrés sur son disque dur ainsi que celle des sites pour lesquels il a demandé de mémoriser sa décision (autorisation ou refus systématique) de placement de cookies. Il peut également supprimer de manière sélective des cookies ou des sites de ces deux listes.

Remarquons que, malheureusement, la configuration par défaut de la grande majorité des navigateurs (pour ne pas dire tous !) est réglée pour permettre à tous les sites d'utiliser (placer et/ou lire) des cookies sans avertissement ni distinction de leur provenance ; l'internaute doit donc modifier explicitement ses préférences pour changer cette situation à moins qu'il ne considère pas les cookies comme une invasion de la vie privée et n'est pas préoccupé de connaître qui en place sur sa machine, dans ce cas il gardera les options inchangées. Récemment, la configuration de certains navigateurs récents peut être réglée de manière à ce qu'ils autorisent ou non le placement des cookies selon leur provenance : l'internaute peut désormais accepter que les sites consultés placent leurs propres cookies, tout en bloquant automatiquement les cookies tiers. C'est une option intéressante pour les internautes inquiétés par les implications en matière de confidentialité des cookies tiers mais pas par les cookies provenant du site visité proprement dit.

II.2.3. La mémoire cache

Nous allons décrire maintenant les limites de la mémoire cache, qu'elle soit située sur l'ordinateur d'un internaute ou dans le disque dur d'un proxyserveur. Rappelons que la mémoire cache du browser a pour fonction de permettre aux sites de se charger plus rapidement s'ils sont revisités ; par conséquent, effacer son contenu peut impliquer un léger ralentissement de la navigation.

a) Sur l'ordinateur d'un internaute

Un internaute peut toujours avoir un minimum de contrôle sur le logiciel qu'il utilise pour surfer sur Internet. Comme on l'a constaté dans la section précédente, les browsers les plus répandus disposent tous d'une fonction de mémoire cache qu'il est possible de paramétrer mais aussi de désactiver (dans la configuration par défaut, elle est activée), mais pas toujours. Mais il est possible qu'une telle fonctionnalité ne soit pas présente sur certains navigateurs, c'est-à-dire que le browser ne crée pas et ne se sert pas de fichiers cache, ou que cette fonctionnalité ne soit pas exploitée par le browser dans sa configuration par défaut.

Si cette fonctionnalité est présente mais désactivée, aucun fichier cache ne sera enregistré sur le disque dur de l'ordinateur de l'internaute. Si par contre le navigateur ne prévoit pas que cette fonction soit désactivée, l'enregistrement en cache des informations concernant la session de l'internaute a lieu et le seul moyen qu'il reste est de trouver le répertoire où sont stockées ces informations. A la fin de sa session lorsqu'il s'est déconnecté, l'internaute n'a plus qu'à effacer manuellement le contenu de ce répertoire qui contient les fichiers cache. S'il oublie d'effectuer cette opération, les fichiers resteront sur son disque dur (sachant qu'ils peuvent prendre une place plus ou moins grande).

b) Sur un serveur proxy

A moins de connaître l'utilisation qu'un fournisseur d'accès fait de son proxy (par exemple, si ce dernier a une politique transparente et en informe ses clients), rien ne permet à

un internaute de savoir si l'administrateur du proxy a mis en place un fichier d'audit. Si c'est le cas, il peut lui aussi conserver une trace exacte des documents téléchargés par l'internaute, cela va des pages HTML visionnées jusqu'aux programmes téléchargés.

Cependant, il est facile de contourner un proxy qui pourrait effectuer le traçage de ses utilisateurs : si pour une raison ou pour une autre, un internaute ne désire plus que les pages des sites qu'il consulte soient stockées (même provisoirement) sur le proxy de son fournisseur d'accès, il a la possibilité de configurer son navigateur pour qu'il effectue une connexion directe à Internet. Pour cela, il lui suffit de supprimer, dans les paramétrages du navigateur, les informations relatives à ce proxy. En effet, un proxy n'est autre qu'un serveur et le navigateur doit disposer de son adresse IP pour pouvoir dialoguer avec lui. Cette adresse a donc nécessairement été enregistrée dans le navigateur afin que ses requêtes lui soient envoyées à lui plutôt qu'au serveur destinataire directement. Si l'on supprime ces informations, ou si l'on indique au navigateur que l'on ne désire plus utiliser ce proxy, les requêtes envoyées par le browser iront directement aux serveurs à qui elles sont destinées. Ceci ne rendra pas impossible tout traçage des connexions par le fournisseur d'accès mais au moins il ne pourra simplement plus utiliser son proxyserveur à cette fin. En contrepartie, les requêtes envoyées par l'internaute ne seront plus optimisées, ce qui peut provoquer sans doute une baisse dans les performances de sa connexion. Toutefois, il faut faire la part des choses puisque la perte de célérité de la connexion sera surtout plus importante pour les sites Web les plus populaires (en général, les gros portails), qui ont plus de chance d'être présents dans leur globalité sur les disques durs des proxies des fournisseurs d'accès. Par contre, on gagnera du temps pour la grande majorité des autres sites parce qu'à chaque requête envoyée, il n'y a plus de délai lié à la recherche précédemment effectuée sur le proxy pour déterminer si la page désirée y était présente ou non.

Pour être complet, signalons que les navigateurs de dernière génération permettent, tout en utilisant le proxy d'un fournisseur d'accès (ou d'une entreprise), d'indiquer une liste de sites pour lesquels les requêtes ne passeront pas par le proxy. Ainsi, cette fonctionnalité peut permettre aux internautes un minimum de discrétion sur certains sites qu'ils considèrent comme sensibles.

II.2.4. Java et ActiveX

De nouvelles évolutions logicielles permettent aujourd'hui de faire beaucoup plus de choses à partir d'un serveur Web que d'afficher des écrans d'information. Le langage Java, tout comme les ActiveX, permet d'exécuter de véritables applications interactives originales. Or ces nouvelles techniques intègrent aussi la possibilité de manipuler des fichiers sur l'ordinateur client. Même si la sécurité a été prise en compte et que, pour Java comme pour ActiveX, l'utilisateur a la possibilité de fixer lui-même le niveau de sécurité qu'il souhaite, encore faut-il qu'il soit conscient et informé des tenants et aboutissants de ces techniques ainsi que des réglages, souvent nombreux et variés, qui lui sont offerts dans certains navigateurs pour personnaliser ce niveau de sécurité. On doit certainement exiger de la part des éditeurs de logiciels et des administrateurs de sites Web le plus d'informations possible, mais cela ne doit pas dispenser l'internaute de faire lui-même un effort d'information.

Tout système informatique peut contenir des failles de sécurité qu'un programmeur habile peut exploiter. De telles failles de conception s'expliquent à la fois par la complexité croissante des logiciels et par la nécessité pour leurs producteurs de les lancer sur le marché très rapidement, risquant ainsi d'oublier certains tests. Les administrateurs de systèmes

informatiques se sont cependant organisés pour lutter contre l'exploitation de telles failles de sécurité[75]. Ce qu'il convient de retenir, c'est que lorsqu'une faille est découverte et exploitée par un individu, elle ne tarde généralement pas à être découverte également par d'autres acteurs d'Internet, notamment les administrateurs des serveurs Web. Ils peuvent ainsi prendre les mesures nécessaires pour faire cesser cette nuisance, tout en informant la communauté des internautes.

Si certaines de ces failles ont beaucoup fait parler d'elles, force est de constater qu'il n'a pas toujours fallu beaucoup de temps pour les rendre inopérantes. Tout dépend de la rapidité d'action des éditeurs de logiciels pour mettre au point un patch correctif. Ainsi, il est probable que le danger que ces menaces représentent, s'il existe toujours, n'est pas véritablement un risque durable pour un internaute à condition qu'il veille à mettre à jour régulièrement ses logiciels, sans oublier que chaque jour, il en apparaît de nouvelles. D'où la nécessité pour tout un chacun de se munir d'un antivirus récent et efficace, dont les définitions de virus sont mises à jour toutes les semaines, et d'un firewall software installé sur son ordinateur afin de se protéger des attaques extérieures[76].

II.3 Logiciels et Services d'« anonymisation »

La première section nous a montré que la configuration par défaut des navigateurs Internet n'assure pas un niveau optimal de sécurité. Ainsi, les cookies sont le plus souvent acceptés sans distinction, le scripting (JavaScript et VBScript) est activé, les applets Java et les composants ActiveX sont exécutés. La deuxième section nous a permis de nous rendre compte des moyens qui nous sont offerts dans le paramétrage des navigateurs pour modifier cette configuration par défaut afin de nous assurer un niveau de sécurité plus important. Dans cette section, nous allons faire un tour d'horizon plus approfondi sur les intermédiaires du réseau et les logiciels clients qui permettent d'améliorer encore ce niveau de sécurité.

II.3.1. Serveurs proxies : un autre point de vue

Tous les serveurs proxies n'appartiennent pas forcément à des fournisseurs d'accès. Ainsi, il en existe, privés ou publics, qui sont utilisables, gratuitement ou non, par tout internaute soucieux de son anonymat. Parmi eux, peu sont ceux qui sont réellement non transparents (càd anonymes, par abus de langage). Mais grâce à ceux-là, on peut s'assurer d'une certaine confidentialité vis-à-vis de son fournisseur d'accès qui ignorera tout du parcours sur le Web de cet internaute puisque chaque requête a lieu sur le proxy, qui se chargera ensuite de le connecter sur le site désiré. Bien sûr, certains serveurs proxies enregistreront son passage dans leurs fichiers d'audit mais de toute façon, l'internaute est sûr d'être anonyme vis-à-vis des sites Web qu'il visite (à condition qu'il n'y ait pas un cookie identifiant de ce site sur son ordinateur, qui lui soit envoyé par son browser).
Il arrive aussi que certains sites Web tiennent à jour une liste « noire » d'adresses IP de proxies. Dans ce cas-là, le site pourrait refuser la connexion et il faudra donc tenter d'utiliser les services d'un autre proxy.

---

[75] Notamment grâce au CERT de l'Université de Carnegie Melon (http://www.cert.org/).

[76] Nous en reparlerons davantage dans la section n°4, mais remarquons d'ores et déjà qu'il en existe de très bons disponibles gratuitement sur le net.

Comment exploiter ces serveurs proxies à partir de son navigateur ? Le plus souvent, ils s'utilisent en plaçant une certaine ligne de commande dans son navigateur (dans la section « Paramètres du réseau local » d'Internet Explorer, par exemple). Une telle ligne de commande contiendra un paramètre du type *http://www.proxyserveur_privé_ou_public.com:n°port*[77] indiquant au browser l'adresse et le port du serveur proxy auquel il doit adresser dorénavant toutes ses requêtes HTTP. La plupart des sites Web anonymiseurs proposent cette possibilité d'intégrer, dans le navigateur, une ligne de commande vers leur propre serveur proxy. Avantage : plus besoin d'aller sur le site Web offrant ce service pour en bénéficier puisqu'une fois configurer de façon adéquate, il suffit d'utiliser son navigateur normalement pour surfer anonymement.

Ainsi, il existe sur Internet des serveurs proxies publics et anonymes (càd qu'ils sont non transparents), accessibles gratuitement et dans certaines circonstances[78] à tous (sans login ni mot de passe). On peut en trouver des listes, mises à jour tous les mois environ, à l'une des adresses suivantes : http://www.multiproxy.org/anon_list.htm,
http://tools.rosinstrument.com/proxy/,
http://directory.google.com/Top/Computers/Internet/Proxies/Free/?tc=1/,
http://netspy.ukrpack.net/index.shtml,
http://poisk.hypermart.net/proxy.htm,
http://www.bugbrother.com/www.cameleon.org/index-2.html#proxytoday.
Pour les utiliser, il faut recourir le plus souvent à la ligne de commande à intégrer dans son navigateur.

Comme les proxies peuvent effectuer un audit de leurs utilisateurs, l'idéal est d'en chaîner plusieurs simultanément, de préférence localisés dans des pays différents. Les risques de remonter jusqu'à l'utilisateur initial sont alors quasi-nuls. Autre possibilité, réalisable en pratique grâce au logiciel ***MultiProxy***[79] (logiciel gratuit) ou ***Proxy Spy***[80], est de changer régulièrement et aléatoirement de serveur proxy, ce qui brouille les pistes. C'est assez efficace puisque même si chaque proxy fait des fichiers d'audit, on ne peut pas être identifié tout le

---

[77] En fait, les numéros de port (un port donné correspond habituellement à un service, ou une application spécifique ; chaque ordinateur connecté à Internet dispose au total de 65 536 ports différents) le plus souvent réservés à la communication avec un serveur proxy sont les 80, 3128, 8000 et 8080.

[78] En effet, quelquefois ils appartiennent à des entreprises ou des administrations, et comme ils sont censés être fermés au public, on comprend alors pourquoi certains d'entre eux ne fonctionnent qu'à certains moments de la journée, ou à certaines périodes.

[25] MultiProxy permet de gérer des listes de proxies (transparents ou non) et d'en choisir un ou plusieurs (alternativement) pour la navigation (anonyme, à condition que le proxy choisi soit non transparent) sur le Web. En bref, il est capable de :

- charger facilement une liste de proxies depuis un simple fichier texte,
- tester la vitesse de chacun de ceux-ci (au moyen du « ping »),
- tester leur degré de « non transparence » (ne fonctionne pas toujours correctement),
- les classer par ordre de rapidité,
- permettre d'utiliser les meilleurs proxies anonymes (ou non) pour la navigation, et éventuellement de les alterner régulièrement selon le désir de son utilisateur.

Son principal point fort : il permet de changer de proxy automatiquement, à chaque requête (séquentiellement ou aléatoirement).

[80] HTTP Proxy-Spy permet d'établir une liste de 20 proxies externes et de commuter rapidement de l'un à l'autre. Il peut également ignorer les requêtes à destination de sites prédéfinis pour filtrer des bannières, par exemple.

long de sa navigation, mais le défaut majeur reste le délai parfois très important entre une requête donnée et sa réponse (les proxies gratuits sont souvent terriblement lents). Néanmoins, avec un peu de persévérance, il est possible de trouver des proxies qui soient suffisamment rapides et « sûrs » pour être utilisables dans de bonnes conditions. Si on en utilise plusieurs à la fois, et de manière aléatoire, alors on peut jouir d'un niveau d'anonymat tout à fait appréciable. Malheureusement, l'anonymat n'est jamais absolument parfait avec ce genre de techniques de masquage d'adresses IP (en anglais, « IP masquerade ») car certains sites peuvent avoir recours à des scripts (JavaScript) ou à des mini-applications (applets Java, ActiveX) pour déterminer la véritable adresse IP d'un visiteur quel que soit le nombre de proxy derrière lesquels il est caché (sauf bien sûr si un de ces proxy est pourvu de fonctions supplémentaires comme ceux de certains sites Web anonymiseurs qui permettent de bloquer le scripting côté client). D'où l'importance de désactiver ses fonctionnalités dans son navigateur si on veut être « tranquille » !

II.3.2. Logiciels de protection des données à caractère personnel

Comme nous l'avons vu ci-dessus, les serveurs proxy permettent d'utiliser d'autres ordinateurs connectés à Internet comme des intermédiaires. Les proxies admettent typiquement les requêtes de pages Web qui leur sont envoyées, demandent les pages requises et transfèrent les résultats à leurs destinataires. Ainsi quand on utilise un proxy, le site auquel on accède profilera les activités du proxy plutôt que les nôtres. Plusieurs produits décrits ci-dessous utilisent les proxies de différentes manières. Le résultat final, cependant, est le même – les activités en ligne d'un utilisateur sont protégées.

Quelques applications qui fonctionnent en tâche de fond sur l'ordinateur de l'internaute :

- ***Anonymity4Proxy*** (shareware, $45) : Conçu par la société iNetPrivacy Software Inc. (http://www.inetprivacy.com), ce logiciel fonctionne comme un serveur proxy local[81] et fait appel à un large réseau de serveurs proxy non transparents dans le monde entier pour protéger l'identité de son utilisateur. N'importe quelle requête que celui-ci émet pour visiter un site sur Internet est d'abord envoyée à travers un de ces serveurs proxy. Ce serveur achèvera la requête et renverra le site Web demandé à l'utilisateur (HTTP, HTTPS et FTP sont supportés), tandis que le site qui a été accédé ignorera le véritable initiateur de la requête. Anonymity 4 Proxy permet d'accéder à une base de données de plusieurs centaines de serveurs proxy anonymes. Il fournit des informations précises à propos de chacun d'entre eux (vitesse de connexion, anonymat, temps de réponse,...) et aide son utilisateur à choisir le proxy le plus rapide en fonction de ses besoins (pour un site donné, selon sa localisation géographique, ou différemment pour chaque requête émise). Des fonctions avancées permettent également de cacher activement la présence de son utilisateur sur Internet, notamment en générant une fausse adresse IP pour chaque requête, en bloquant ou en modifiant n'importe quelle information envoyée par le browser (cookies, bavardage,...) et en simulant des requêtes ordinaires (en cachant les différentes traces laissées par l'activité d'un proxy, comme si la requête allait directement au site Web voulu). L'internaute peut partager

[81] Par abus de langage, un proxy local est une application s'exécutant en local, càd sur l'ordinateur de l'internaute, et se comportant d'une certaine manière comme un serveur proxy vis-à-vis de l'internaute (avec ou sans antémémoire et le plus souvent non transparent pour permettre l'anonymat de celui-ci, mais rien n'empêche qu'elle soit également dotée d'autres fonctions, comme le filtrage des cookies ou du bavardage du navigateur par exemple, afin de protéger davantage les données à caractère personnel de son utilisateur) si ce n'est qu'en principe, seul l'internaute lui-même peut accéder à ses services (ce n'est donc pas un serveur, au sens habituel du terme).

une connexion anonyme sur Internet avec d'autres personnes appartenant au même réseau local ou utiliser des listes noires pour bloquer les sites qu'il désire éviter. Le programme offre un large éventail d'options techniques afin que son utilisateur puisse mieux contrôler la diffusion de ses informations en ligne (empêchant le traçage par un site au moyen des fichiers d'audit) et l'utilisation des multiples serveurs proxies disponibles au moyen de règles personnalisées. Sa gestion centralisée évite ainsi d'avoir besoin de (re)paramètrer sans cesse son navigateur. C'est l'un des produits les plus anciens et les plus réputés. Mais le principal défaut d'A4Proxy est qu'il reste payant tandis qu'à l'heure actuelle, d'autres logiciels du même type, aussi efficaces, sont disponibles gratuitement sur le Web.

- ***JavaAnonProxy*** (JAP – http://anon.inf.tu-dresden.de/index_en.html, logiciel gratuit) est un outil très efficace pour la protection de la vie privée sur Internet. D'après ces créateurs, « *L'utilisation des services du Net à l'heure actuelle est aussi la source des traces informatiques que l'on laisse sur son passage. De plus en plus de compagnies essaient d'utiliser ces traces pour créer des profils individuels des internautes. Nous aidons les gens à protéger leur « e-Privacy »*[82] *: le logiciel JAP fournit une communication anonyme et non observable sur le Net. Grâce à cette base n'importe quel « e-service » apparenté à la vie privée peut être utilisé plus sereinement. Même dans le domaine du e-commerce, la protection de la vie privée joue un rôle important. La plupart des gens n'apprécient pas de recevoir des e-mails non sollicités (du « Spam ») comme résultat de leurs activités sur Internet.*

Objectif : la navigation sur le Web sans laisser de traces. Nous offrons un logiciel open source facile à utiliser pour tous : JAP aide à protéger la vie privée de ses utilisateurs et couvre tout type de trafic sur le réseau Internet. Certains des systèmes existants de communication anonyme réduisent la sécurité au profit de meilleure performance. Notre système fournit un anonymat fort, càd que le système résiste à l'analyse du trafic. Cela signifie que même un adversaire qui voudrait écouter toutes les communications sur Internet ou les liens sur le réseau n'obtiendrait aucune information à propos de qui communique avec qui.

Qu'est-ce que JAP ? Le système JAP consiste en un logiciel client (appelé JAP, pour JavaAnonProxy) et d'une chaîne de plusieurs serveurs intermédiaires. Voici les deux scénarios les plus communs lors de l'utilisation de JAP :

- *JAP aide à protéger la vie intime d'un internaute. JAP peut être installé sur l'ordinateur d'un utilisateur pour protéger ses activités sur Internet. C'est très utile pour un usage d'Internet pour un particulier.*
- *JAP aide à protéger une organisation. JAP peut aussi être installé sur une machine dédiée, par exemple un firewall ou un serveur proxy standard. JAP servira alors comme une entrée protégée (au niveau de la confidentialité) pour la compagnie toute entière, et il n'y a plus besoin d'installer le logiciel sur chacune des stations de travail. Ceci peut être très utile pour des sociétés afin de dissimuler leurs transactions et/ou activités de recherche sur Internet, pour contrer les observations de compétiteurs ou d'autres activités d'espionnage industriel.*

Comment fonctionne-t-il ? JAP agit comme un proxy local entre le browser et l'Internet insécurisé. Toutes les requêtes pour les pages Web sont traitées par JAP et sont chiffrées plusieurs fois. Les messages chiffrés sont envoyés à travers une chaîne de serveurs

[82] NDA : c'est-à-dire leur vie privée sur Internet (qui est loin d'être « virtuelle » bien entendu...)

intermédiaires (nommés « Mixes » par l'inventeur de cette théorie, David Chaum) à la destination finale sur Internet.
Les multiples couches de chiffrement protègent tous les messages. Un Mix collecte les messages dans un groupe, change totalement leur apparence (enlève une couche de chiffrement) et les réexpédie tous en même temps, mais dans un ordre différent. Un adversaire peut observer tous les liens de communication, cependant il ne peut pas déterminer une relation entre les paquets qui entrent et ceux qui sortent. Ainsi un internaute reste anonyme à l'intérieur du groupe de tous les utilisateurs du service.
De façon démontrable, le système protège efficacement la vie privée de ses utilisateurs tant que le Mix travaille correctement. Malheureusement personne ne sait si un certain Mix est en réalité fiable ou non. C'est pourquoi nous utilisons une chaîne complète de Mixes. Chaque message passe au travers de plusieurs Mixes et la chaîne entière de Mixes doit être corrompue pour observer avec succès les activités d'un utilisateur. Le chaînage empêche l'emploi d'un seul Mix pour l'observation des communications. C'est cela la signification d'anonymat fort : même le service d'anonymat lui-même ne peut pas espionner ses utilisateurs. »

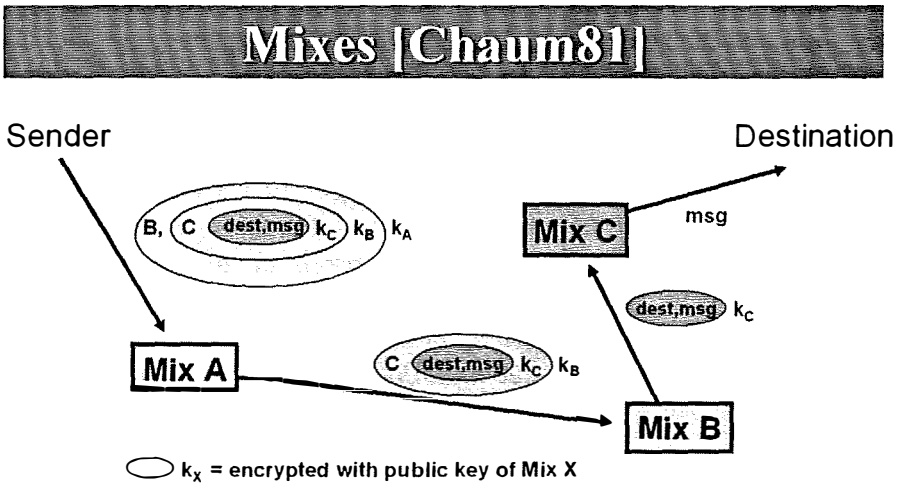

Figure II-1 (source : référence [13] page 22)

- ***Internet Junkbuster Proxy*** (http://www.junkbusters.com/, logiciel gratuit) : Agissant comme un proxy local, il est capable de « neutraliser » le bavardage des browsers en le modifiant astucieusement, de faire disparaître du browser, pendant la navigation, les bannières publicitaires et de « manger » les cookies afin de protéger la vie privée de ses utilisateurs. Junkbuster bloque les requêtes concernant les URLs qui concordent avec un « blockfile » paramétrable soit défini par son utilisateur soit fourni par d'autres utilisateurs de Junkbuster. Le logiciel bloque aussi les cookies non autorisés (en se basant sur une liste des domaines de cookie acceptés ; le refus étant l'option par défaut) et modifie le bavardage du navigateur. L'utilisateur peut configurer à sa guise ces deux précédentes fonctions. Ainsi, il peut choisir d'accepter les cookies venant d'un site mais faire en sorte que ces cookies ne puissent plus être envoyés au site qui les a émis (ou inversement, il peut accepter que les cookies présents sur son disque dur soient envoyés aux sites qui les ont émis mais empêcher ces sites de redéposer de nouveau cookie ou de modifier un cookie existant). L'utilisateur peut entre autre

indiquer aux sites Web visités une fausse signature de navigateur ou une fausse adresse de page référente.
Le programme base sa décision de permettre ou non à des fichiers d'atteindre un ordinateur où il est installé sur base du contenu du blockfile utilisé. Ce blockfile contient une liste des URLs (ou plutôt d'ensembles d'URLs) dont l'utilisateur ne souhaite pas recevoir de fichiers, le plus souvent les cookies ou les annonces publicitaires qui accompagnent les sites Web que l'internaute désire visiter. Le blockfile étant paramétrable, on peut ajouter ou supprimer des URLs ou même changer de blockfile à n'importe quel instant pour le remplacer par un autre. De vastes communautés d'internautes postent leurs blockfiles sur le Web pour qu'ils soient utilisés comme un guide des sites qui devraient être bloqués parce qu'ils recueillent des informations personnelles.
Nous conseillons également le logiciel gratuit ***Guidescope***[83] (http://www.guidescope.com/), soutenu par les auteurs de Junkbuster. Il existe un projet de système de filtrage écrit en Java, appelé ***Muffin*** (http://muffin.doit.org/) sous licence GPL (et donc open source), incluant plusieurs filtres qui peuvent éliminer les cookies, stopper les animations GIF, supprimer les bannières publicitaires, ajouter/enlever/modifier les tags HTML arbitraires, ôter les applets Java et les JavaScripts, réécrire les URLs, modifier le bavardage des navigateurs, et plus encore. Il supporte les protocoles HTTP et HTTPS (SSL). Les utilisateurs peuvent créer leurs propres filtres en Java grâce à une API fournie.

- ***Anonymizer*** (Gratuit pour le service de base, $10 pour 3 mois ($30 pour une année) avec le Service Premium), créé par Anonymizer.com (http://www.anonymizer.com) :
Anonymizer, un des plus ancieux services de ce type sur Internet, agit comme un intermédiaire entre un internaute et les sites qu'il visite, dissimulant son adresse IP aux propriétaires des sites en question. Pour utiliser ce service, outre l'accès à leur site Web ou à leur proxy au moyen de son navigateur, comme évoqué précédemment, l'internaute peut également télécharger sur le site Web de la société un petit logiciel, appelé Anonymizer Privacy Toolbar, se greffant sur son browser (plug-in) et qui lui permet de surfer anonymement (et de bénéficier également de différentes protections supplémentaires pour préserver sa vie privée, la version gratuite possédant moins de fonctionnalités que la version payante) de manière tout à fait transparente sans devoir modifier la configuration de son browser. C'est certainement le service le plus simple à utiliser puisqu'il possède une interface facile à maîtriser, mais en contre partie, il ne dispose que de quelques possibilités de personnalisation.

- ***Naviscope*** (logiciel gratuit), publié par Naviscope (http://www.naviscope.com) :
Naviscope est un plug-in pour le navigateur qui simultanément protège la vie privée en ligne et accélère la navigation sur le Web. Le logiciel Naviscope prend place entre le browser et l'Internet. Il va chercher les informations que son utilisateur demande et les lui retourne. L'utilisateur peut choisir de ne pas permettre la livraison de cookies, de bannières publicitaires, ou d'autres objets Web comme les sons ou les arrière-plans, mais aussi de cacher les champs Referer et User Agent dans le bavardage du navigateur, d'empêcher les fenêtres pop-up et/ou l'exécution de codes JavaScript. Chaque outil de filtrage peut être administré globalement ou être paramétré différemment selon le site. Ce programme accélère également la navigation grâce à plusieurs techniques dont une connue sous le nom de « prefetching ». Parce que Naviscope agit aussi comme un proxy local, il peut rester actif après qu'il ait délivré la page Web qui a été demandée. Basé selon la requête, Naviscope

[83] Ce logiciel est similaire à Webwasher (voir ci-dessous). Il permet de bloquer les bannières publicitaires, les Web bugs et les cookies, d'accélérer la navigation, de modifier le bavardage du navigateur. En outre, il offre un certain nombre d'options diverses et utiles.

utilise des mots clés pour télécharger des pages Web que probablement l'internaute visionnera prochainement. Si, par exemple l'internaute effectue une recherche sur Google correspondant aux mots clés « Vie Privée », Naviscope téléchargera les pages des prochains résultats de la recherche pendant que l'internaute regardera le premier lien. En pré-chargeant ce contenu additionnel, Naviscope facilite une vision plus rapide des pages Web, toutes les pages pré-chargées bénéficiant également des « améliorations » en matière de protection de la vie privée. L'option de « prefetching » est bien sûr entièrement personnalisable. Mais l'accélération de la navigation se fait aussi, au choix, en créant une mémoire cache des résolutions DNS, en effectuant des connexions persistantes avec les serveurs Web et/ou en optimisant certains paramètres des protocoles TCP/IP. Très complet et facile à utiliser, Naviscope est un excellent compagnon, qui dispose de surcroît d'un panneau de contrôle discret indiquant de nombreuses informations sur la connexion Internet et comportant plusieurs fonctions pratiques supplémentaires.

- ***Webwasher*** (http://www.webwasher.com, gratuit pour un usage non commercial) est un filtre Internet développé à l'origine par Siemens. Il est installé, en complément du navigateur comme plug-in, sur l'ordinateur de son utilisateur et permet d'accélérer la navigation. Il filtre toutes les requêtes HTTP ainsi que les pages Web avant qu'elles n'apparaissent à l'écran pour supprimer au choix bannières publicitaires, fenêtres pop-up, cookies (les cookies peuvent aussi être classés en différentes catégories qui correspondent aux différentes méthodes de filtrage à utiliser pour les traiter), images animées, Web bugs, JavaScript, applets Java, bavardage du navigateur, etc. Ainsi, ce logiciel empêche les sites tiers de réaliser des profils personnels de ses utilisateurs, de les pister et de suivre les traces qu'ils laisseraient sur leur passage s'ils ne l'utilisaient pas mais aussi de connaître les mots-clés entrés sur un moteur de recherche. Les cookies non sollicités ne peuvent être ni envoyés ni reçus.
Ce logiciel est utilisé comme proxy local et peut être partagé par plusieurs utilisateurs.
Une aide pas à pas (en anglais) guide l'utilisateur dans la configuration du logiciel. Tous les filtres sont personnalisables et peuvent être activés/désactivés indépendamment les uns des autres. Ainsi, par exemple celui pour fermer automatiquement les fenêtres pop-up peut être configuré selon des critères de taille, de type et/ou d'URLs. WebWasher gère également les connexions à un proxy distant et les domaines à consulter sans proxy.
En fait, tout ou presque est configurable selon les envies de son utilisateur dans ce nettoyeur de pub aux multiples fonctions. Jusqu'à la possibilité de créer une liste d'URLs agréées. Hors de cette liste pas de possibilité d'accès (en anglais « white list »). Il existe aussi la « black list », càd la liste des sites interdits.
Un équivalent shareware tout aussi complet est ***SpyBlocker*** (http://www.spyblocker-software.com/spyblocker/).

- ***Proxomitron*** (site officiel : http://proxomitron.org/, traduction française http://go.to/proxomitron-fr) est un outil gratuit de contrôle des informations qui proviennent du Web. Grâce à lui, l'utilisateur peut choisir ce qu'il souhaite voir arriver sur son écran. C'est ainsi que les bannières publicitaires, les fenêtres pop-up, les fichiers multimédia (sons et vidéos) sont susceptibles d'être contrôlés, voire tout simplement éliminés, sans oublier les textes qui défilent dans la barre d'état du navigateur et les images qui s'affichent en arrière-plan. Ce logiciel fonctionne comme un proxy local. Comme toutes les requêtes envoyées par le navigateur Web et les réponses des serveurs des sites Web fréquentés passent par lui, il peut les filtrer au passage. Lorsque la phase de configuration du navigateur est terminée, il suffit de cocher des cases pour filtrer les pages Web en conséquence. Proxomitron sait geler les animations GIF, masquer l'identité du navigateur, bloquer l'utilisation des JavaScripts,

désactiver « l'anti-clic bouton droit de la souris », supprimer les « frames » des pages Web ainsi que gérer les cookies. Il permet encore de ne désactiver qu'un certain nombre de commandes JavaScript en laissant les autres disponibles. Mais les fonctionnalités de ce logiciel ne s'arrêtent pas là, puisqu'il est capable également d'appliquer des règles de filtrage vis-à-vis des serveurs Web. Ces règles, l'utilisateur peut les éditer lui-même de façon à exercer un contrôle maximum sur le contenu des informations échangées par son navigateur. Dans cette perspective, Proxomitron propose même la possibilité d'appliquer des filtres particuliers à des pages Web spécifiques. Certes, l'utilisateur peut éprouver une certaine difficulté à appréhender l'ensemble des possibilités offertes mais globalement, Proxomitron est facile à mettre en œuvre pour un utilisateur averti. Il bénéficie en outre d'une aide complète. La liste de tous les filtres (tous pouvant être largement paramétrés, on peut même éditer ses propres filtres) proposés par Proxomitron prouve qu'il est sans doute un des outils de filtrage les plus complets disponibles actuellement.

- ***Ultimate Anonymity*** (http://www.ultimate-anonymity.com/) est un service commercial ($15) procurant tout le nécessaire pour naviguer sur Internet anonymement, envoyer du courrier électronique chiffré, poster et lire anonymement des messages sur les newsgroups et converser sur IRC/ICQ ou les « Chat » sans dévoiler sa véritable adresse IP. Idéal pour les novices.

II.3.3. Les services d'anonymat en ligne

Une nouvelle catégorie de services chargés de protéger la vie privée de leurs utilisateurs a fait son apparition depuis quelque temps sur Internet, leur principal avantage : l'efficacité. Pour offrir un niveau de sécurité parmi les plus élevés, ils associent souvent une application tournant en tâche de fond à un serveur proxy en ligne. Ces formules (appelées mixtes lorsqu'elles conjuguent à la fois logiciel et serveur proxy distant) permettent ainsi à leurs utilisateurs de surfer anonymement sur des sites Web tout en prévenant tout profilage par un site tiers. Pour ce faire, leurs connexions passent d'abord par un serveur d'anonymisation qui cache leur adresse IP mais les protège également contre toute utilisation illicite de leurs données personnelles ou toute attaque pouvant survenir sur le réseau en agissant comme un bouclier entre leur ordinateur et les différentes menaces (principalement les plus courantes) qui pourraient les atteindre en ligne que ce soit du point de vue de la sécurité ou de la vie privée.

Le principal problème lié à l'utilisation de ce genre de services est que l'internaute doit avoir une grande confiance vis-à-vis de la société qui gère le système d'anonymisation et édite le logiciel, car rien n'empêche cette société d'être au courant de tout ce que l'utilisateur fait sur le Web.

En général, ces systèmes proposent différents services qui correspondent aux différentes tâches nécessaires pour garantir l'efficacité de la protection :

- réécrire les pages Web que l'utilisateur veut consulter, avant de les lui envoyer, à partir de leurs serveurs protégés pour empêcher le profilage effectué par certains sites et améliorer la sécurité de son ordinateur face aux multiples menaces cachées venant du Web.

- cacher l'adresse IP unique (au moins, lors d'une session donnée) de l'ordinateur de l'utilisateur lors de ses déplacements d'un site à un autre mais aussi vis-à-vis de tous les autres serveurs extérieurs, ce qui les empêchent de pister l'utilisateur

(celui-ci devenant « invisible »), ou mieux chiffrer l'ensemble des informations transmises avec Internet.

- empêcher les parties extérieures de mettre des fichiers malveillants sur le disque dur de l'utilisateur, d'utiliser un code malicieux pour porter atteinte à l'intégrité de son ordinateur ou de créer des fichiers d'audit traçant précisément sa navigation.

- neutraliser les cookies, le bavardage des navigateurs (càd la communication des variables d'environnement), le scripting côté client (scripts, applets Java et ActiveX) potentiellement malicieux, nuisible ou carrément dangereux, les Web bugs et quelques autres risques vis-à-vis de la vie privée de l'internaute.

Certains systèmes d'anonymisation permettant l'utilisation de certaines techniques de cryptographie, ils apportent un niveau supplémentaire de protection du contenu des échanges, pour garantir leur intégrité, leur confidentialité et pour authentifier les interlocuteurs.
Bien entendu, ces services sont rarement gratuits, et l'internaute doit souscrire à un abonnement mensuel ou annuel pour y avoir accès et bénéficier de leurs avantages.

Le serveur d'anonymisation utilisé conjointement avec le logiciel n'est, en général, rien d'autre qu'un proxy « amélioré », càd doté des ingrédients supplémentaires pour garantir une protection efficace contre les codes ou les scripts menaçants et pour effectuer les filtrages nécessaires. Faisant les requêtes en son nom, il intercepte donc cookies, applets Java et JavaScripts/VBScripts ainsi que les ActiveX et les bandeaux publicitaires, évitant ainsi à l'internaute d'être identifié et profilé, puis il lui renvoie les pages « anonymisées » en léger différé (à peine perceptible). C'est donc en passant par ce serveur avant de se connecter à d'autres serveurs d'Internet qu'un utilisateur est assuré que ses données à caractère personnel ne sont pas communiquées aux sites Web qu'il désire visiter par la suite. Le serveur d'anonymisation est ainsi capable de protéger ses utilisateurs sans pour autant causer des effets secondaires nuisant gravement à la navigation interactive à l'intérieur d'un site Web, si ce n'est le léger délai à chaque requête/réponse (délai dû au traitement par le serveur de toutes les informations échangées). Cependant, il arrive que certains services augmentent la temporisation de la redirection des pages vers le navigateur en contrepartie de la gratuité.

Ces technologies sont souvent basées sur ce que l'on nomme un protecteur d'identité. Un protecteur d'identité peut être considéré comme un élément du système qui contrôle la divulgation de la véritable identité d'un individu à différents processus au sein du système d'anonymisation. Une des fonctions les plus importantes de ce protecteur d'identité est de convertir l'identité réelle d'un utilisateur en une pseudo-identité, identité alternative électronique, que l'utilisateur peut adopter lorsqu'il utilise le système.
Plusieurs techniques peuvent être utilisées pour introduire un protecteur d'identité dans un tel système; notamment les techniques de cryptographie impliquant des signatures électroniques ou des pseudonymes électroniques.

Des exemples de systèmes complets d'anonymisation sont fournis par de nombreuses compagnies : nous en exposons quelques-uns ci-dessous.
Ainsi, Anonymizer (http://www.anonymizer.com) offre les services suivants, regroupé sous le nom d'***Anonymizer Private Surfing***, à ces utilisateurs payants[84] (Utilisateur Premium) :

- Protection contre le profilage en ligne et le spamming : ces services permettent en toute sûreté d'accepter les cookies des sites qui les requièrent (pour faire du shopping, accéder à un contenu personnalisé ou s'identifier) sans être inquiété par un pistage à long terme. Ainsi, Anonymiser converti les cookies ayant une date d'expiration lointaine et qui sont envoyés à ses utilisateurs en cookies de session, qui expire automatiquement une fois que la connexion est terminée. Mais ces services empêchent aussi les sociétés de cybermarketing et les spammeurs de déposer des cookies sur l'ordinateur d'un utilisateur, enlèvent la plupart des bannières publicitaires se trouvant dans les pages Web visitées, éliminent les Web bugs et bloquent les fenêtres pop-up, ce qui accélère également la navigation (le contenu « inutile » n'étant plus téléchargé).
- Messagerie & Newsgroups anonymes : ce service permet d'envoyer, à partir du site Web de Anonymiser, des e-mails privés qui resteront anonymes et impossibles à tracer. Il offre également la possibilité de consulter des forums de discussion, les « chats », messagerie instantanée IRC/ICQ, en tout anonymat.
- Défense contre les menaces du Web : ces services permettent d'éliminer le contenu actif dangereux (scripting côté client dommageable) en le réécrivant, de modifier le bavardage du navigateur (empêchant la transmission de la page référente ainsi que l'identification du browser et du système d'exploitation utilisés) et de masquer l'adresse IP de l'utilisateur (en reroutant les requêtes et réponses HTTP par les serveurs d'Anonymizer) pour une navigation anonyme et sûre (protection contre toute personne ou organisation malintentionnée).

Les services offrent en outre une protection spécialisée pour le trafic HTTPS ou le streaming multimédia. Ils sont en outre facile d'utilisation et ne nécessitent aucun réglage compliqué. En outre, ils sont accessibles quel que soit l'endroit où l'on se connecte à Internet, puisqu'ils ne nécessitent qu'un petit logiciel à télécharger, dont on peut même se passer pour accéder aux services, et sont personnalisables grâce à plusieurs niveaux de sécurité et de protection de la vie privée paramétrables.

Anonymizer offre donc essentiellement deux services, le surf (ainsi que le transfert de fichiers) protégé et l'e-mail anonyme. Néanmoins, il y a tout de même une petite faille dans ce genre de système. En effet, comme l'internaute doit se connecter au serveur distant pour ensuite bénéficier des services d'anonymisation, la connexion entre son ordinateur et ce serveur reste vulnérable à une surveillance par un tiers, à moins qu'elle ne soit protégée. C'est un service supplémentaire de Anonymizer :

- Sécurité pour toutes les activités du Web : ces services chiffrent les requêtes émises pour l'obtention de pages Web, ainsi que le contenu des pages demandées lors de leur

[84] Par rapport au service payant, Anonymizer propose un service de base qui est complètement gratuit. Mais le service Premium reste plus rapide et plus confortable à utiliser puisqu'il est jusqu'à 4 fois plus rapide que le service gratuit (qui est moins « prioritaire », d'où des délais d'attente supplémentaires dans les communications). De plus, il ne dispose d'aucune limitation dans le nombre de pages visionnées, d'aucune restriction ni dans la taille des fichiers téléchargés, ni dans l'utilisation de streaming multimédia. Enfin, il est le seul à permettre l'emploi du protocole HTTPS, à enlever les pop-ups gênants et les annonces publicitaires encombrantes. Néanmoins, le service gratuit permet de bénéficier également du blocage des cookies, du JavaScript et des applets Java sans toutefois avoir la possibilité de contrôler plus finement ce filtrage. Signalons que le ralentissement du chargement des pages Web est intentionnel pour ce service gratuit.

retour aux utilisateurs, de telle façon que leur fournisseur d'accès, ou toute autre tierce personne qui pourrait observer leur connexion Internet, ne puisse créer un fichier d'audit. Pour cela, les services correspondants utilisent le protocole SSL pour les échanges entre l'ordinateur de l'utilisateur et les serveurs d'Anonymizer (préservant la confidentialité des échanges et détectant les fraudes ou altérations éventuelles), mais chiffrent également les URLs et les titres des pages visitées (aucune tierce personne ne peut alors déterminer l'historique des sites consultés, gardant secret les habitudes et intérêts de navigation de l'utilisateur). Cela permet aussi d'effectuer des achats en ligne avec une sécurité accrue.

Un ensemble de services similaires à ceux proposés par Anonymizer sont ceux des sociétés suivantes :

- IDzap (http://www.idzap.com/), dont le service est intitulé ***IDsecure Anonymous Browsing Services*** (gratuit pour un service limité, $15 pour 3 mois ($50 pour l'année) avec le service complet),
- Orangatango (http://www.orangatango.com/) et son service appelé ***VirtualBrowser*** (7 jours d'essai gratuit, $40 pour une année d'utilisation),
- Somebody Inc. (http://www.somebody.net/) et son produit ***Somebody v4.0***[85] ($129 par mois pour le service complet).

Les internautes sont régulièrement sollicités à communiquer des informations personnelles à des sites Web lors de leur navigation dans le but d'accéder à un contenu particularisé. Avec chaque demande d'informations, toutefois, il y a certains choix à faire : il faut se décider entre donner des informations liées au métier que l'on exerce ou des informations privées, choisir combien d'informations on est prêt à fournir à chaque site, choisir si on donne au site des informations vraies ou non (dans le cas d'informations fausses, on parlera généralement d'informations « anonymes »), et enfin, il faut fréquemment choisir un nom d'utilisateur (un « user ID ») et un mot de passe à fournir au site.
Les possibilités varient d'un site à l'autre bien entendu, et peuvent être déconcertantes pour des consommateurs. Certains produits ont pour but de simplifier ce processus d'échange d'informations, tout en fournissant en même temps des protections plus efficaces à l'utilisateur pour préserver sa vie privée. Les produits suivants, Freedom, Persona et Privada, en sont des exemples mais il ne faut pas le confondre avec les services de gestion de l'identité tels que le « Passport » de Microsoft ou « DigitalMe » de Novell. Si ces services peuvent s'avérer pratiques, c'est contre l'échange d'une partie de la vie privée de leurs utilisateurs, ces sociétés cherchant moins à les servir qu'à engranger un maximum de données personnelles à des fins marketing.

L'éditeur canadien Zero-Knowledge (http://www.freedom.net) propose une application appelée ***Freedom WebSecure***. Ce programme permet de protéger ses messages (par e-mail ou chat) et de surfer anonymement. Il permet donc d'éviter de laisser des traces utilisables par les marchands de publicité sur le Web ou d'autres spammeurs éventuels, mais surtout de défendre sa vie privée. Et cela grâce à des protocoles et des algorithmes de chiffrement réputés forts,

[85] Somebody 4.0 est un service utilisant un proxy distant pour founir un accès anonyme à Internet. Ce proxy cache l'identité d'un utilisateur et empêche que des informations personnelles soient révélées par son navigateur, les cookies ou du scripting côté client lors de sa navigation. Il masque également son adresse IP lorsqu'il utilise IRC ou ICQ. De plus, ce service permet à l'utilisateur d'effectuer des appels internationaux en utilisant des produits de téléphonie pour Internet. Il fournit un service DNS ainsi qu'un service de courrier électronique anonyme, permettant à l'utilisateur d'envoyer des e-mails sans divulguer son identité. Le proxy supporte les protocoles suivants : HTTP, HTTPS, FTP, Gopher et WAIS.

cette solution se basant sur au moins trois relais TCP/IP combinés à des algorithmes de chiffrage puissants (au moins 128 bits) basés sur un cryptosystème asymétrique. Ainsi, comme les protocoles TCP/IP sont employés par tous les services du Net, ils sont donc tous chiffrés et rendus anonymes de la même façon. Freedom étend donc ses capacités à presque toutes les utilisations d'Internet : navigation Web, messageries (e-mail et Chat), groupes de discussions (Forums) et service Telnet. Chacun des trois relais intermédiaires par lesquels les requêtes transitent ne connaît que l'adresse IP de son prédécesseur[86]. La société garantit qu'ils ne conservent aucun fichier-registre, de telle sorte que même deux relais mis ensemble sont incapables de retracer les informations recherchées ou importées. Le routage de l'information est évidemment dynamique et susceptible d'être modifié même au cours d'une communication brève. Un système de gestion et filtrage des cookies est aussi intégré dans Freedom. Il protége la vie privée de ses utilisateurs en utilisant des « nyms » (sorte de pseudonymes) totalement indépendants de leur identité réelle et est aussi capable de filtrer les mails non sollicités (« spamming »). En outre, il bloque les annonces publicitaires, notamment pour accélérer la navigation et améliorer la protection de la vie privée. Enfin, il neutralise les contenus actifs potentiellement dangereux (scripting côté client) et bloque l'exécution de programmes malicieux afin de préserver l'internaute des menaces éventuelles compormettant la sécurité de son ordinateur et/ou sa vie privée.
Le logiciel, qui réside sur l'ordinateur d'un utilisateur, permet la création et la gestion de multiples pseudonymes, qui correspondront à la création et à la maintenance de différentes personnalités persistantes qui reflètent divers niveaux d'anonymat. Utilisant un ou plusieurs « nyms », les utilisateurs peuvent visiter n'importe quel site, envoyer des e-mails, poster des messages dans des newsgroups et même utiliser IRC sans craindre que leur adresse IP réelle ou leur adresse e-mail ne soient démasquées. En effet, lorsque l'utilisateur commence ses activités en ligne au moyen de ses pseudonymes, Freedom les chiffre grâce à plusieurs niveaux de cryptographie – ceci fait qu'il est impossible, en principe, de profiler son ordinateur par l'intermédiaire de son pseudonyme. Le système offre une couche supplémentaire de protection en routant tout le trafic au travers de nombreux serveurs formant « Freedom Network ». Ce réseau apporte une série de chemins détournés améliorant significativement l'anonymat. L'application peut également se lier directement à n'importe quel logiciel de messagerie POP3 pour protéger le contenu des messages. Le système chiffre les courriers électroniques lorsqu'ils voyagent dans le réseau Freedom et remplace l'adresse de retour par une adresse de la forme « nym@freedom.net », les réponses seront alors routées vers la boîte aux lettres d'origine de façon totalement transparente pour l'utilisateur.
Zero-Knowledge a également inclut des protections contre de nombreux e-mails commerciaux, qui peuvent constituer du Spam pour l'utilisateur. Le Freedom Network rejette automatiquement les utilisateurs qui envoient des quantités excessives d'e-mails à partir de leurs pseudonymes. Alternativement, n'importe qui peut aller sur le site Web de Freedom et refuser la réception d'e-mails en provenance de certains pseudonymes.
Zero-Knowledge a fait beaucoup d'effort pour rendre Freedom plus agréable à utiliser. L'application, qui administre les nyms et les connexions à Internet au travers du réseau Freedom, requière un minimum de connaissances techniques. Néanmoins, la plupart des utilisateurs, même ceux possédant quelques aptitudes de base dans la technologie d'Internet,

[86] En fait, même s'il y a chiffrage, les adresses IP (de l'origine et de la destination) contenues dans les paquets HTTP envoyés restent toujours en clair (normal sinon la communication ne serait pas possible !), d'où la nécessité d'utiliser un ou plusieurs serveurs proxies non transparents pour cacher complètement l'origine des paquets à la destination ou à un tiers. Signalons aussi que le chiffrage prend un certain temps (nécessaire aux calculs effectués par les algorithmes correspondants, qui consomment des ressources système plus ou moins importantes selon la puissance du chiffrement), d'où une certaine lenteur dans les communications qui peut se faire sentir (la confidentialité a un prix !).

devraient être capables de régler Freedom et de le faire tourner avec succès. Freedom semble être conçu avant tout pour travailler sur une machine autonome et non à l'intérieur d'un réseau d'entreprise. L'installation prend quelques minutes, pendant lesquelles les nom, pseudonyme et e-mail de l'utilisateur sont intimement liés à un numéro de série généré de manière aléatoire. Une phrase clé est demandée pour s'identifier ou ouvrir des messages chiffrés. Selon la société, avis recoupé par d'autres experts, seul ce numéro de série est détenu par Zero Knowledge. Retrouver le nom d'utilisateur à partir de ce numéro serait mathématiquement impossible.

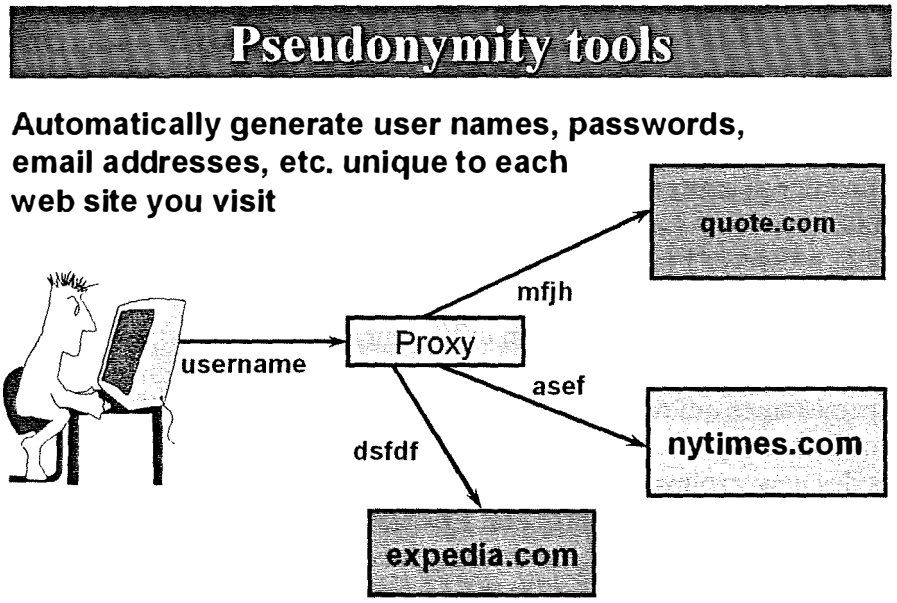

Figure II-2 (source : référence [13] page 15)

Mais l'originalité de ce programme, conçu par le service de Ian Goldberg (cfr. [17]), expert reconnu de la cryptographie, se situe également ailleurs : aucune information nominative n'est détenue par Zero Knowledge, l'éditeur lui-même est donc incapable de lever le voile sur l'identité de ses clients ! D'éventuelles poursuites judiciaires resteraient impuissantes.
Jusqu'à présent les nombreux sites dédiés à l'anonymat sur le Web avaient été contraints de fermer leurs portes les uns après les autres à la suite de plaintes ou de procédures judiciaires les obligeant à lever le voile sur l'anonymat de certains de leurs clients.
Freedom, disponible uniquement en version PC pour MS Windows, peut être commandé en ligne sur le site de Zero Knowledge. La version commerciale coûte environ 59.95$ pour une année avec cinq pseudonymes complets. C'est un programme qui tourne en tâche de fond et dont on peut modifier le niveau de sécurité à tout moment avec la contrepartie d'une lenteur croissante de connexion pour les niveaux de sécurité les plus exigeants puisque le principe, en plus du cryptage des données personnelles de l'utilisateur, est le reroutage des paquets d'information par un, deux voire trois serveurs en fonction du niveau de sécurité souhaité. Ce qui a pour effet de masquer la véritable adresse IP de l'utilisateur.
D'autres services en ligne offrant des capacités d'anonymat et de confidentialité existent déjà sur le net, mais, à la différence de Zero Knowledge, ces services peuvent être poussés, en cas de mandat de perquisition, à révéler l'identité d'un compte anonyme pour en retrouver l'auteur. Freedom, conforme pour être vendu selon les termes de la loi canadienne, devient donc une arme à double tranchant : d'un côté, il permet à un citoyen, par exemple militant politique qui se sent menacé, d'exercer son droit à la confidentialité (reconnu par la déclaration des droits

de l'homme) mais, de l'autre, il peut devenir complice d'actes de malveillance ou de harcèlement électronique sans que le coupable puisse être inquiété. L'éditeur affirme toutefois qu'un éventuel comportement illégal de la part d'un client peut entraîner son exclusion.
Tout ceci fait de Freedom sans doute le meilleur et le plus complet service d'anonymat disponible.

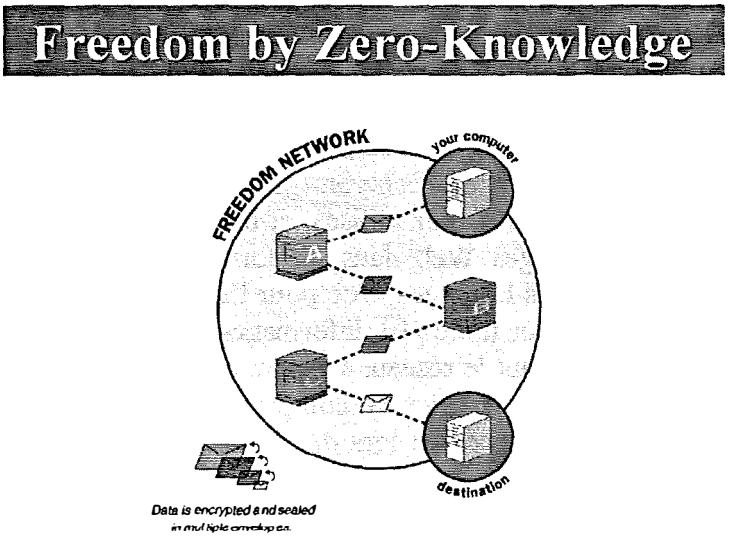

Figure II-3 (source : référence [13] page 23)

Persona/Child Persona sont deux produits gratuits publiés par PrivaSeek (http://www.privaseek.com) et permettant de gérer son « identité » en ligne.
Ainsi, Privaseek a mis au point deux produits liés qui ont pour fonction d'administrer les informations personnelles de leurs utilisateurs lorsqu'ils accèdent à Internet. Leurs profils sont protégés grâce à l'emploi d'un « Persona ». Cela ne coûte rien de créer son propre Persona. Une fois que l'on est enregistré dans le service, un petit fichier, appelé le « PersonaAgent », est installé sur notre ordinateur. Ce PersonaAgent protège toutes nos informations personnelles au moyen d'un mot de passe, nous permet de remplir quasi instantanément les formulaires Web, et nous donne la capacité d'administrer rapidement et facilement notre identité en ligne. Quand on s'inscrit comme un Persona, on a le choix de télécharger un petit programme appelé « PersonaValet », une barre à outils qui améliorent la navigation sur le Web, qui permet de gérer ses préférences en matière de vie privée, qui permet de surveiller les cookies reçus (et de les contrôler) et qui dispose en sus d'une option qui nous alerte lorsqu'un site a une politique P3P valide (cette option permet alors d'effectuer la comparaison de cette politique avec nos préférences en matière de vie privée).
Privaseek a également créé une version de leur programme Persona spécifiquement adaptée pour les enfants, appelé « Child Persona ». Avec cet outil, les parents sont capables de notifier les sites Web de la présence d'un enfant de moins de 13 ans. Le programme verrouille aussi les informations personnelles de l'enfant dans un « coffre virtuel », empêchant que ces informations deviennent accessibles aux sites Web participants. En participant à ce programme, les sites Web reçoivent une garantie supplémentaire qu'ils sont en conformité avec le « Children's Online Privacy Protection Act » de la législation américaine, publié en 1998.

Un autre exemple de service mixte est donné par l'entreprise californienne Privada (http://www.privada.com). Cette société offre des services appuyant tous les types de transactions, y inclus la navigation, l'e-mail, l'envoi de messages et, bientôt, le e-commerce. L'infrastructure de Privada est basée sur la compartimentalisation et la cryptographie.
PrivadaControl ($5 par mois) est une application qui réside sur l'ordinateur de son utilisateur et lui offre un contrôle de sa vie privée sur Internet. L'application a deux composants majeurs, « Web Incognito » et « Messaging Incognito ». Chacun est paramétrable selon les préférences de l'utilisateur en matière de sécurité.
WebIncognito est un logiciel facile d'emploi permettant de surfer anonymement et sans laisser de trace. Il s'agit en fait d'une version grand public de PrivadaProxy, un produit davantage destiné aux entreprises. Utilisant un proxy et un logiciel de chiffrement installé sur le PC de l'internaute, WebIncognito empêche les sites visités de récupérer l'adresse e-mail, l'adresse IP, le type d'OS et de navigateur utilisés, et bloque les cookies. Il opère en routant toutes les requêtes pour des pages Web dans un format chiffré à l'intérieur du Privada Network. Le réseau retrouve alors les pages Web pour l'utilisateur qui les a demandées. Web Incognito a pour objet de protéger toutes les informations et données de l'utilisateur, dès le point de transaction et à travers tout le réseau, assurant ainsi la protection de la vie privée de l'internaute vis-à-vis de toutes les parties, y compris Privada et son fournisseur d'accès. Chaque cookie que l'utilisateur rencontre lors de sa navigation avec le réseau Privada est attribué à son profil sur le réseau Privada plutôt qu'à lui personnellement, lui donnant les bénéfices de la navigation Web personnalisée sans les soucis liés au respect de sa vie privée. Si un utilisateur rencontre un site digne de confiance et qu'il désire lui faire part de ses informations personnelles, il peut facilement mettre hors d'usage la protection de sa confidentialité due à Web Incognito (et accepter les cookies de ce site). Cette application fonctionne donc comme un coupe-feu pour la vie privée de l'utilisateur. Pour l'employer, l'utilisateur crée un ou plusieurs comptes privés virtuels qui représente ses activités en ligne, joue le rôle de pseudonymes pour l'identité réelle et désassocie complètement toutes ses informations personnelles de ses faits et gestes. Web Incognito permet à son utilisateur de créer ou de détruire ses identités électroniques, de choisir entre elles lorsqu'il interagit en ligne et de gérer ses propres attributs et caractéristiques. Le système permet de bloquer rapidement les cookies et d'autres contenus actifs, comme les contrôles ActiveX et les applets Java, qui peuvent être potentiellement dangereux pour la vie privée. Mais si l'utilisateur ne souhaite pas bloquer systématiquement tous les applets Java, cookies ou contrôles ActiveX, Web Incognito lui permet de décider du niveau auquel la personnalisation et les services Web peuvent fonctionner. Les cookies sont placés sur des serveurs centralisés au sein du réseau Privada, et non sur l'ordinateur de l'utilisateur. Tous les fichiers journaux ou tentatives d'extraction de données sont associés à l'identité en ligne de l'utilisateur et non à son identité réelle. Privada soutient que les utilisateurs peuvent aisément éliminer un ou tous les cookies qui auraient été installés. Malheureusement, ce service est seulement disponible aux États-Unis. De plus, il ralentit assez fortement la vitesse des communications.
Messaging Incognito étend les avantages de la protection de la vie privée aux activités liées au courrier électronique. Quand un utilisateur envoie un e-mail au moyen de son compte Privada, le message est chiffré, signé digitalement et transmis par le Privada Network. Le réseau assigne alors au message un délai aléatoire de 30 minutes à 4 heures, ce qui empêche une tierce partie de déterminer son identité à partir du moment où le message a été émis. Ce délai peut aisément être désactivé pour certains messages marqués « haute priorité ».

Crowds est un système expérimental mis au point par AT&T Research (http://www.research.att.com/projects/crowds/, cfr. [19]). L'idée de base de ce système est novatrice puisqu'il s'agit du premier système capable de « cacher » les transferts de données sur le Web sans nécessiter la confiance d'une autorité centrale. Il permet de cacher les adresses IP de ses utilisateurs vis-à-vis des serveurs finaux (ceux hébergeant les sites Web visités) mais aussi des autres membres du système (dont nous détaillerons brièvement le fonctionnement infra), des administrateurs de ce système et encore de n'importe quelle tierce personne « écoutant » les communications sur le réseau.
Le principe de ce système repose sur une large communauté d'internautes, appelée « crowd » (la foule), répartie dans le monde, et sur l'idée d'un « mélange dans la foule » (càd cacher les actions d'une personne parmi celles de beaucoup d'autres). Lorsqu'un internaute veut masquer son adresse IP, il rejoint la communauté des autres utilisateurs. Ensuite, dès qu'un membre envoie une requête HTTP vers un serveur Web quelconque, celle-ci est envoyée soit directement au serveur soit à un autre membre choisi aléatoirement. Ce membre peut alors choisir, en toute indépendance, de la transmettre au serveur final ou de la renvoyer à un autre membre choisi aléatoirement. Et ainsi de suite jusqu'à ce que la requête atteigne finalement sa destination. Ainsi, cette requête peut « voyager » parmi les ordinateurs de plusieurs autres membres, tous choisis aléatoirement, qui la relaient à tour de rôle. Après un certain nombre de « rebonds » au sein de la communauté, la requête est envoyée vers le serveur destinataire pour y être traitée. La réponse revient en faisant le chemin inverse. Le serveur final est donc incapable de connaître l'origine d'une requête qu'il a reçue puisqu'elle peut venir de n'importe quel membre de la communauté. Il en est de même pour les membres du Crowd eux-mêmes, car un membre qui envoie une requête à un autre l'a peut-être lui-même reçue d'un autre membre, et ainsi de suite (le nombre de rebonds déjà effectués n'étant pas connu par les membres puisqu'il est arbitraire).

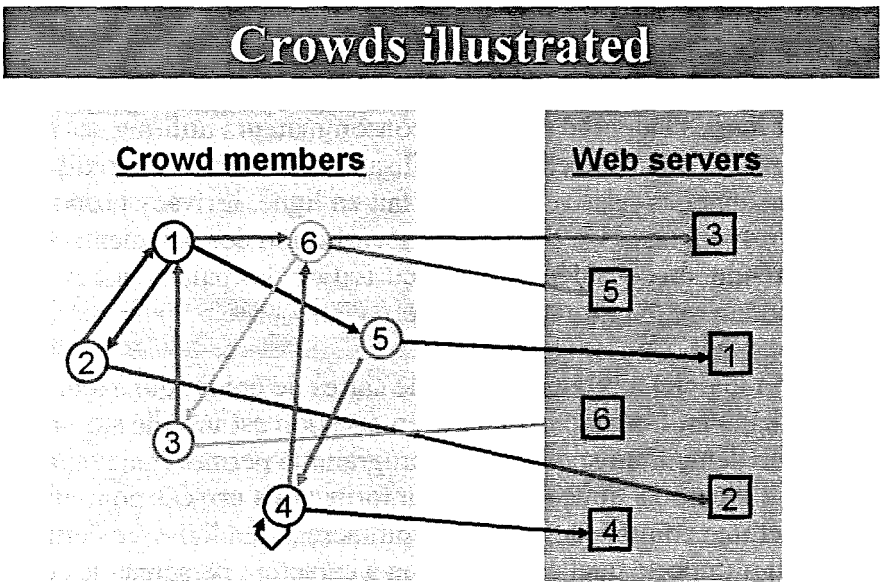

Figure II-4 (source : référence [13] page 26)

Pour rejoindre un Crowd, il suffit d'installer le proxy mis au point par AT&T et de configurer son browser pour qu'il lui transmette ses requêtes.
Quels sont les risques ? Si nous sommes membre d'un Crowd, notre ordinateur peut transmettre à des serveurs Web des requêtes qu'il n'a pas lui-même initiée, puisque le proxy fonctionnant sur notre machine enverra des requêtes provenant d'autres membres. Si un

hacker malintentionné profite du Crowd pour mener à bien ses attaques sans être identifié, il est possible qu'on nous accuse à sa place. Dans ce cas, il n'y a pas d'autre alternative que d'expliquer que l'on est membre d'un Crowd et que par conséquent on ne contrôle pas les requêtes issues de notre ordinateur. Un second risque est que, à cause du routage de nos requêtes à travers plusieurs ordinateurs, le système peut accroître le risque de divulgation des données contenues dans les requêtes et les réponses correspondantes. Ainsi, bien qu'un membre d'un Crowd ne peut déterminer l'origine (la destination) d'une requêtre (d'une réponse) donnée, il est tout de même capable d'accéder à son contenu. Ce problème peut prendre toute son ampleur lorsque le contenu d'une requête contient par exemple un mot de passe confidentiel pour accéder à un site particulier. Dans un tel cas où un site a besoin de nous authentifier, le Crowd ne protège plus tes données à caractère personnel et il est alors recommandé de quitter le Crowd (en modifiant la configuration du navigateur) lorsque l'on désire consulter un tel site.

Il a été souvent argumenté que même si l'on peut être capable de naviguer sur Internet anonymement, notre vie privée doit finalement être livrée si l'on désire faire l'achat de biens ou de services en ligne. Jusqu'à un certain point, on doit donner des informations personnelles, incluant les informations d'expédition et le sensible numéro de la carte de crédit, aux commerçants du Web.
Les produits suivants permettent de changer la donne en ce qui concerne les achats privés en ligne. Au moyen de l'utilisation d'une méthode complexe de commerce électronique, il y a maintenant la possibilité de participer au marché du e-commerce sans livrer ses informations personnelles.

Le système ***iPrivacy*** qui est proposé par iPrivacy LLC (http://www.iprivacy.com, prix et disponibilité inconnus, cfr. [18]) permet, selon ses concepteurs, un commerce électronique anonyme, depuis la visite de sites jusqu'à la livraison. Il permet aux utilisateurs de son logiciel de naviguer et de chercher des produits sur le Net, de les acheter et d'arranger leur livraison de manière confidentielle, sans jamais révéler l'identité de l'acheteur. Selon la société, elle-même ne connaîtrait jamais l'identité réelle des consommateurs utilisant ses services. En ce qui concerne la transaction, seuls le client et l'utilisateur de la carte de crédit auraient accès aux informations personnelles relatives à l'achat fait en ligne. iPrivacy proposent donc des offres qui associent respect de la vie privée et la sécurisation des paiements. Il s'agit ici d'une nouvelle façon de prendre en considération la sécurisation des paiements. Ainsi, le numéro de carte de crédit devient au même titre que le numéro de Sécurité Sociale une donnée nominative dont on entend maîtriser la diffusion.
iPrivacy distribue son logiciel aux fournisseurs de cartes de crédit, qui à leur tour le fournissent aux particuliers. Lorsque le logiciel en question est installé sur un ordinateur, la navigation sur le Web demeurera anonyme. Le programme permet également à un utilisateur d'établir un compte « identité » qui contient ses informations privées pour effectuer des achats. Quand il émet une commande avec une commerçant en ligne, ce dernier ne voit pas son identité réelle ou n'importe quelle information à caractère personnel le concernant. Les informations qu'il inscrit sont déléguées à son compte identité, qui est attaché à un compte existant dans sa banque. Le prix est rapporté sur le relevé de sa carte de crédit – comme si l'achat avait été effectué dans le monde réel.
La protection de la vie privée d'un utilisateur préserve l'entièreté du processus d'achat, mais elle est aussi étendue à l'entièreté du processus d'expédition. Ainsi quand une transaction est complète, le produit est rattaché à une adresse réelle seulement lors du processus de livraison au moment où le paquet est étiqueté.

En résumé, les clients envoient l'ensemble des informations relatives à leur achat en ligne de façon cryptée. Ensuite, chaque acteur de la commande prend connaissance uniquement des informations qui lui sont nécessaires. Seul le commerçant dispose des informations relatives au bien acheté, par exemple le titre du livre. Le transporteur est le seul à disposer de l'adresse de livraison. Il sait qu'il s'agit d'un livre mais il en ignore le titre. Enfin, la banque est la seule à recevoir les informations relatives au moyen de paiement mais elle ne connaît pas la nature du bien acheté.
Comment iPrivacy entend-elle convaincre les fournisseurs de cartes de crédit d'utiliser ses produits ? Simplement en leur expliquant que l'achat anonyme est un service supplémentaire qu'ils peuvent rendre à leurs clients, c'est donc un nouveau facteur de différenciation face à la concurrence.

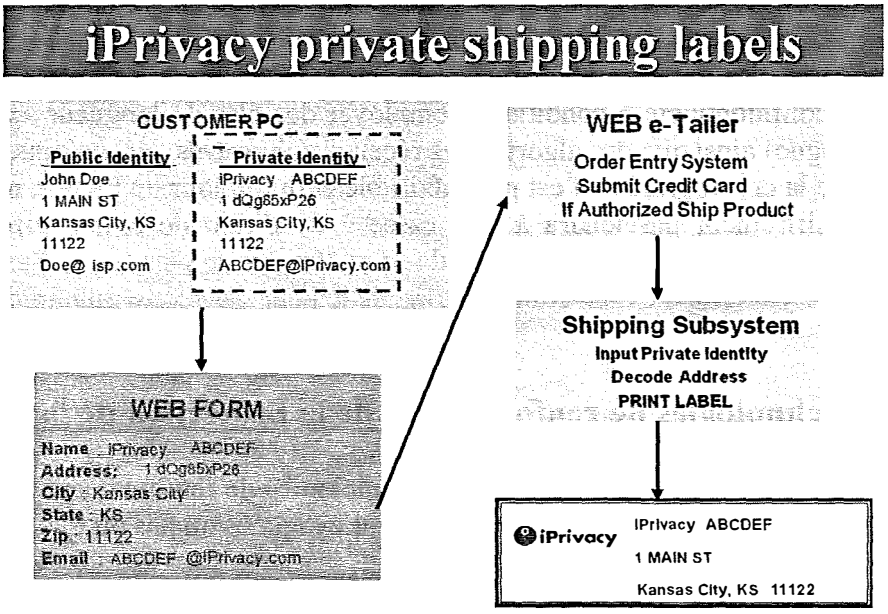

Figure II-5 (source : référence [13] page 16)

Enfin, ***ZixCharge*** est un logiciel gratuit publié par ZixIt (http://www.zixcharge.com). ZixCharge est un système d'autorisation de transactions qui permet à ses utilisateurs d'acheter des biens sur Internet tout en préservant du même coup leur confidentialité. Il utilise un système spécial pour préserver leurs informations à caractère personnel des mains des commerçants sur le Web.
Après s'être enregistré auprès du service ZixCharge, on obtient un « portefeuille » virtuel pour le Web. On peut utiliser plusieurs méthodes de paiement électronique, à choisir selon nos préférences, pour « remplir » ce portefeuille, celles-ci incluent bien sûr la carte de crédit mais ce n'est pas la seule. Les commerçants participant au programme ZixCharge ne réclament plus les informations pour l'expédition et pour le financement. A la place, ils demandent simplement confirmation à ZixCharge que l'utilisateur est bien autorisé à faire ses achats. ZixCharge propose à ses utilisateurs des « factures » virtuelles qui détaillent les achats qu'ils sont en train de faire. Une fois qu'un utilisateur confirme qu'il souhaite acheter les éléments listés, ZixCharge transmet l'achat au compte approprié et masque sa véritable identité en ligne. Ainsi le seul groupe qui possède ces informations à n'importe quel moment est seulement ZixCharge. La seule raison pour laquelle ZixCharge révèlerait les informations personnelles d'un utilisateur aux e-commerçants est dans l'éventualité d'une contestation de la transaction.

Le logiciel donne aussi à ses utilisateurs la capacité de contrôler les e-mails qu'ils pourraient recevoir de la part des commerçants visités. Ainsi, si un utilisateur ne désire plus recevoir de courriers électroniques en provenance d'un certain commerçant, il lui suffit de le noter dans son profil et ZixCharge ne transmettra plus les messages correspondants à son destinataire.

Bien souvent, la cryptographie est présentée comme l'outil idéal et sans faille de la protection des données[87]. Mais il faut savoir que son usage reste limité car, en général, seul le contenu des échanges est chiffré, par contre les adresses des interlocuteurs circulent en clair sur le réseau. A l'heure actuelle, il existe sur Internet une multitude d'outils de chiffrage. Or le grand public est loin de posséder les connaissances suffisantes pour choisir un algorithme de chiffrement adéquat. De nombreuses méthodes de chiffrage considérées comme « inviolables » il y a quelque temps à peine sont aujourd'hui « cassées » en quelques heures ou jours, les moyens et outils informatiques évoluant en puissance et en performance chaque mois. Ainsi, la cryptographie pourrait constituer un moyen fiable garantissant un niveau élevé de confidentialité à condition d'employer des clés de longueur appropriée (càd suffisamment longue) ainsi que des algorithmes récents et éprouvés. Néanmoins, il ne faut pas perdre de vue que la cryptographie n'est pas infaillible et qu'un jour ou l'autre, n'importe quel algorithme de chiffrement parviendra à être cassé à condition de réunir suffisamment de puissance de calcul (cfr. [4]). Il importe donc de se mettre à jour sans cesse. Sans oublier que l'utilisation de la cryptographie a un coût en terme de lenteur dans les communications.

II.4. Autres technologies de renforcement de la protection de la vie privée

Cette section décrit et analyse certaines technologies de renforcement de la protection de la vie privée qui n'ont été encore que partiellement évoquées ou pas du tout.

II.4.1. Les logiciels tueurs de cookies et nettoyeurs de fichiers résiduels

Tandis que les cookies peuvent avoir beaucoup d'utilisations avantageuses (les mots de passe dont le site se « souvient » ou permettre le e-commerce « one-click »), ils causent aussi d'importants soucis sur le plan de la vie privée (permettant la création d'un profil de la navigation de l'internaute sur le Web dans un but marketing, ou pire). Deux types de réaction au problème suscité par les cookies pour le respect de la vie privée sont analysés ci-dessous. Le premier trouve son origine dans le secteur Internet lui-même et a été incorporé dans les principaux navigateurs existant sur le marché. Le second provient de divers groupes de défense de la protection de la vie privée et de maisons de logiciels. Il consiste en outils permettant de détruire tous les cookies ou certains d'entre eux.

Sous la pression des associations américaines et européennes de défense de la vie privée, les éditeurs de logiciels ont été contraints d'intégrer des fonctions de désactivation des cookies au sein de leurs navigateurs. Ainsi, le gouvernement américain s'est élevé à plusieurs

[87] En matière de vie privée, il s'agit de s'assurer de la confidentialité des données du point de vue du client. Mais la cryptographie ne se limite pas qu'à cela puisqu'elle constitue un des principaux outils en matière de sécurité des données sur les réseaux de communication, la sécurité des données regroupant l'intégrité, la disponibilité et la confidentialité des données du point de vue client et serveur ainsi que l'authentification des acteurs de la communication.

reprises contre les excès de certains sites qui collectent des données sur leurs visiteurs pour ensuite les submerger de publicité[88].

La seule tentative de résoudre le problème des cookies par le secteur Internet est le mécanisme anti-cookies mis en oeuvre dans certains navigateurs depuis leur version 3 (pour Netscape et Internet Explorer). Dans ces différentes versions, un utilisateur averti d'Internet pouvait paramétrer son navigateur en choisissant entre trois options:

– accepter tous les cookies
– refuser tous les cookies (Netscape et MSIE) ou les cookies qui ne proviennent pas du serveur original (Netscape seulement)
– accepter ou refuser au cas par cas, sur demande

Ces mécanismes anti-cookies étaient cependant insuffisants, et ce pour de nombreuses raisons :

1. habituellement le choix par défaut est celui qui respecte le moins la vie privée (accepter tous les cookies) et l'utilisateur moyen ne sait pas que le cookie est largement utilisé par les sites Web ni dans quel but puisque les sociétés de cybermarketing tracent, par exemple, les mots-clés introduits dans les moteurs de recherche en employant des moyens qui lui sont invisibles.
2. Le mécanisme d'opposition des cookies ne joue que dans un sens, celui de la réception de nouveaux cookies, mais ne prévient pas le renvoi systématique et invisible des cookies déjà enregistrés sur son disque dur. Autrement dit, l'internaute ne sait pas et ne peut s'opposer à ce qu'un cookie préalablement reçu soit renvoyé au site émetteur.
3. les cookies peuvent être de natures très différentes: certains cookies sont utiles et non identifiants (exemple : langue préférée). D'autres sont identifiants, mais peuvent être employés, ou non, conformément aux règles de protection de la vie privée. On peut généralement dire que les cookies de session sont beaucoup moins menaçants pour le respect de la vie privée que les cookies persistants. Refuser tous les cookies pourrait ne pas être à l'avantage de l'utilisateur Internet.
4. comme nous l'avons déjà évoqué, certains sites Web refusent l'accès aux utilisateurs qui n'acceptent pas les cookies.
5. plusieurs sites Web (ou des sites Web reliés par hyperliens invisibles) envoient des séries de cookies, et une approche au cas par cas obligerait l'utilisateur Internet à les refuser les uns après les autres, générant ce que l'on a nommé la "fatigue du click" qui pourrait par dépit pousser l'utilisateur à accepter les cookies une fois pour toutes, pour ne plus être dérangé dans leur navigation.

[88] Une étude réalisée en 1998 par l'université de Georgetown pour le compte de la Federal Trade Commission (FTC), organisme fédéral chargé de veiller au respect des règles commerciales, rapporte que 92 % des sites commerciaux collectent des données personnelles sur leurs visiteurs, 14 % seulement les en avisent et mettent en œuvre des mesures de protection, alors que 88 % des internautes interrogés dans le cadre de cette étude apprécieraient quant à eux « la possibilité de visiter des sites Web en préservant leur anonymat ». Deux des principales organisations américaines de défense des libertés réclament à l'État fédéral de voter une loi spécifique pour protéger la vie privée des internautes. L'Electronic Privacy Information Center (EPIC) et le Center for Democracy and Technology (CDT) remarquent que le principe d'autorégulation prôné par les sociétés commerciales sur le Web ne fonctionne pas. Les données qu'elles collectent sont revendues pour alimenter le marché croissant du marketing direct en ligne. La FTC, chargée aussi de la protection des consommateurs, s'est contentée de réclamer aux gestionnaires de sites commerciaux d'adopter des « lignes de conduite » communes pour informer les internautes sur l'usage qu'ils font de leurs données privées.

6. dans certains cas, le texte avertissant de l'installation d'un cookie parait incomplet, peu clair, et pourrait être trompeur. De plus, le contenu des cookies eux-mêmes demeure généralement inintelligible car codé.
7. lors de l'installation d'un nouveau navigateur, le premier site visité, qui est par défaut celui de l'éditeur du navigateur, peut envoyer sournoisement un cookie avant que l'utilisateur ait l'occasion de désactiver l'acceptation des cookies.
8. le choix d'accepter ou de refuser est mince. Il faudrait, par exemple, que l'utilisateur puisse déterminer une durée de vie limité, càd modifier le cookie.

En juillet 2000, Microsoft annonçait l'introduction de la version bêta d'un système de sécurité pour la prochaine version d'Internet Explorer, permettant une gestion optimisée des cookies du Web. Comme le promettait les premières descriptions, le mécanisme offre plusieurs caractéristiques permettant aux utilisateurs de mieux contrôler les cookies, c'est le système que nous avons décrit dans la première section pour MSIE 6. Ainsi, le navigateur peut désormais faire la différence entre les cookies provenant de l'interlocuteur principal et ceux d'une tierce partie et l'installation par défaut préviendra l'utilisateur lorsqu'un cookie persistant d'une tierce partie est placé. En outre, cette nouvelle fonction permet aux internautes de détruire tous les cookies d'un simple clic et rend les informations relatives à la sécurité et au respect de la vie privée plus aisément accessibles. Le mécanisme de sécurité cependant n'augmente pas le contrôle des consommateurs sur les cookies de l'interlocuteur principal et qui existent sur la plupart des sites Web commerciaux.

C'est pour toutes les raisons citées supra que des dizaines d'éditeurs indépendants ont développé, depuis quelques années, différents logiciels qui permettent à un internaute de gérer ses cookies et par là, de protéger sa vie privée. Ces logiciels analysent les cookies entrant, en tâche de fond et en temps réel et les acceptent ou les rejettent en fonction des paramètres choisis par l'utilisateur.
Toutefois, en l'absence de tels gestionnaires de cookies, nous conseillons de supprimer systématiquement les cookies à la fin de chaque session Internet.
Il existe de nombreux logiciels gratuits ou shareware que chaque internaute peut librement télécharger sur le Net et utiliser sur son ordinateur pour contrôler les cookies qu'il reçoit ou possède déjà (sans le savoir, peut-être).

Voici ci-dessous une brève analyse sur une dizaine de systèmes de contrôle des cookies, qui représentent la gamme de ce genre de produits :

Cookie Web Kit est une suite logicielle publiée par CookieCentral (http://www.cookiecentral.com). Cet ensemble gratuit de logiciels supprime les cookies sur le disque dur d'un internaute chaque fois qu'il démarre son ordinateur. Alternativement, les fichiers peuvent être supprimés à n'importe quel moment en cliquant sur l'icône du programme. Ce logiciel peut également être paramétré pour supprimer les fichiers de la mémoire cache du browser et l'historique des liens visités.

Cookie Pal (Shareware, $15), développé par Kookaburra Software (http://www.kburra.com), permet d'accepter ou de rejeter automatiquement tous les cookies, mais aussi d'être paramétré selon le choix de son utilisateur pour chaque site qu'il visite. Le logiciel inclut un outil qui permet de consulter à chaque instant le contenu de n'importe quel cookie enregistré sur un ordinateur.

Cookie Master (logiciel gratuit), mis au point par Barefoot Productions (http://www.barefootinc.com), est un utilitaire qui trace les cookies dans un navigateur. Le programme s'interface avec le navigateur pour afficher les détails des cookies et permettre le

contrôle de ces cookies. Le logiciel crée un fichier d'audit enregistrant toutes les activités en rapport avec les cookies (réceptions, enregistrements et envois).

Buzof (Shareware, $10), créé par Basta Computing (http://www.basta.com), permet de répondre, fermer ou minimiser automatiquement presque toutes les fenêtres récurrentes, y compris les fenêtres d'acceptation de cookies affichées par le navigateur. Après avoir indiqué au logiciel comment il doit répondre aux demandes de cookies, Buzof scannera périodiquement l'écran et exécutera l'action qu'on lui a dictée dès qu'une fenêtre apparaît.

PGPcookie.cutter (Shareware, $19.95), publié par PGP (http://www.pgp.com), recueille tous les cookies rencontrés par un utilisateur lors de sa navigation sur le Web. Il les garde dans un endroit protégé jusqu'à ce que l'utilisateur décide de les accepter pour échanger des informations avec des sites de confiance ou de les mettre hors d'usage.

NSClean & IEClean (Shareware, $40) sont édités par Privacy Software Corporation (http://www.nsclean.com). Les deux programmes (un pour Netscape Navigator, l'autre pour Internet Explorer) éliminent les cookies du browser correspondant, ainsi que tous les autres fichiers additionnels (mémoire cache, historique,...). Ils permettent à un utilisateur d'utiliser un alias comme nom d'utilisateur, de supprimer ses favoris, d'effacer tout enregistrement de ses activités sur un newsgroup, et plus encore.

Cookie Cop 2 (logiciel gratuit), développé par PCmag (http://www.pcmag.com/), est un utilitaire permettant de contrôler les cookies site par site, de désactiver les fenêtres pop-ups, enlever le champ Referer du bavardage du navigateur, bloquer les bannières publicitaires, convertir les cookies permanents en cookies de session, et plus.

NoCookie (logiciel gratuit), mis au point par 1.0 Technologies (http://www.onepointoh.com), est un système d'administration des cookies conçu spécifiquement pour MacOS. Il permet d'accepter ou non les cookies de Netscape, d'inspecter les informations contenues dans les cookies, et de supprimer un seul cookie ou tous à la fois.

Cookie Terminator (Shareware, $14.95), créé par 4Developers (http://www.4developers.com), a pour fonction de contrôler les cookies stockés sur le disque dur d'un ordinateur, quel que soit le répertoire où ils se trouvent. Son interface est toujours disponible dans la barre des tâches de MS Windows. CookieTerminator permet de spécifier une élimination systématique et automatique des cookies, et indique clairement d'où proviennent les cookies dans le système.

Cache & Cookie Washer, produit publié par Webroot (http://www.webroot.com/ccwasher.htm), permet de procéder à un nettoyage rapide et profond de l'ordinateur hôte après les navigations de son utilisateur. Il supprime tous les cookies (ou seulement ceux que l'on souhaite éliminer ; c'est utile dans le cas où l'on préfère garder les cookies des sites qui enregistrent uniquement nos préférences), purge la mémoire cache et l'historique des liens visités mais aussi les autres traces laissées par le navigateur selon son type[89]. Il peut aussi fonctionner en arrière plan, sans intervention de l'utilisateur. Il peut être lancé soit automatiquement (périodiquement selon un timing paramétré par l'utilisateur, au démarrage ou à l'arrêt du PC, à la fermeture du browser) soit manuellement. Sa facilité d'utilisation permet de gagner du temps en éliminant, en une seule opération, toutes les traces laissées par un navigateur après une session sur Internet.

Cookie Crusher est publié par TheLimit Software (http://www.thelimitsoft.com); c'est un logiciel gratuit. Il automatise le refus des cookies en temps réel et avant qu'ils ne soient enregistrés sur le disque dur, que l'on utilise Netscape ou Internet Explorer. Bien que ces deux browsers proposent, depuis des versions assez anciennes, d'accepter ou de refuser les cookies par le biais d'une boîte de dialogue, Cookie Crusher offre l'avantage de maintenir une liste des sites pour lesquels les cookies seront acceptés ou refusés de façon transparente et sans

[89] Il existe trois versions du programme Cache & Cookie Washer correspondant respectivement aux trois navigateurs que sont AOL, Netscape et Internet Explorer.

intervention de la part de l'internaute. Ainsi lorsqu'un utilisateur atteint un site Web et décide de refuser un cookie, Cookie Crusher refusera en toute discrétion tous les envois suivants de cookies en provenance de ce site, libérant l'utilisateur de cette tâche lassante. Certains cookies ne seront donc pas sauvegardés sur son disque dur et sa vie privée se trouvera ainsi mieux protégée puisque seul l'utilisateur décide des cookies spécifiques qu'il accepte de garder. Cet utilitaire offre également la possibilité d'afficher et d'effacer les cookies (un par un) stockés sur l'ordinateur mais aussi celle de créer des filtres pour rejeter, lors de la navigation, ceux jugés indésirables ou faciliter le choix à la carte en tenant une liste précise de l'ensemble des cookies acceptés ou rejetés afin de ne plus réclamer d'intervention de l'utilisateur à chaque fois qu'ils sont à nouveau reçus ; bien sûr cette liste peut être modifiée par l'utilisateur à tout moment. Il aide aussi l'internaute à déterminer une classification des cookies selon leur fonction en l'informant de l'usage de tel cookie pour du profilage publicitaire, du shopping en ligne, du pistage ou autre. Cookie Crusher est simple à utiliser et inclut une aide en ligne efficace.

Chaque fois qu'un internaute surfe sur Internet, son navigateur télécharge un certain nombre de fichiers qui restent présents sur le disque dur de son ordinateur même lorsque sa session est terminée. Souvent, ces fichiers sont enregistrés dans différents répertoires disparates, ce qui rend la tâche de leur suppression plus difficile pour un néophyte. Or ces fichiers peuvent divulguer certaines informations à caractère personnel, incluant par exemple les sites Web que l'internaute a visités. Mais le navigateur n'est pas la seule application a engendré de telles données, ainsi le système d'exploitation lui-même enregistre un certain nombre d'activités exécutées par son(ses) utilisateur(s) dans différents fichiers « temporaires ». Il en est de même pour certains logiciels qui mémorisent les derniers documents ouverts, ou d'autres choses encore. Tout cela constitue un ensemble des traces révélatrices des multiples activités opérées par une personne Les produits suivants aident leurs utilisateurs à toutes les éliminer facilement (qu'ils s'agissent des divers fichiers de susceptibles de détailler la navigation en ligne mais aussi les copies temporaires des fichiers multimédia, l'historique des sites visités et des adresses entrées, les mots de passe mémorisés, les données des formulaires complétés en ligne, la liste des documents récemment utilisés ou recherchés, et le nom des derniers documents ouverts par certains programmes populaires, comme MS Office, Real Player,...) :

iClean (http://www.aladdinsys.com/iclean/),
evidence eliminator (http://www.evidence-eliminator.com),
Complete Cleanup (http://www.softdd.com/complete/),
Internet Cleanup (http://www.ontrack.com/internetcleanup/),
et bien d'autres.

Ces programmes sont paramétrables à la fois pour les types des fichiers qu'il supprimera et la fréquence avec laquelle il les éliminera. Ils identifient quels fichiers sont supprimables sans danger et quels sont ceux qu'il faut éviter d'éliminer car ils sont vitaux pour le fonctionnement du système d'exploitation. Certains d'entre eux disposent aussi de plusieurs fonctionnalités supplémentaires (nettoyage du registre ou de la corbeille, trouver les raccourcis « cassés », les fichiers obsolètes ou en multiples exemplaires, les fichiers log créés par des applications, supprimer de façon définitive et sûre les fichiers confidentiels en les écrasant plusieurs fois de suite, ...).
Un équivalent gratuit s'appelle ***Internet Sweeper*** (http://www.bmesite.com/).

D'après nous, ces logiciels, très utiles auparavant, deviennent petit à petit obsolètes avec la modernisation des navigateurs qui commencent, dans leur dernière version, à inclure progressivement le même type d'outils de gestion des cookies.

De plus, certaines remarques similaires à celles présentées ci-dessus peuvent aussi s'appliquer dans le cas de ces logiciels :

1. L'internaute doit traiter ses cookies chaque jour, au cas par cas, à cause des natures différentes des cookies et de leur apparition régulière.
2. Dans le cas des logiciels shareware, l'internaute doit payer pour sa propre protection.
3. Tous les mécanismes anti-cookies disponibles ne sont pas faciles d'emploi ni faciles à comprendre pour un internaute moyen, qui doit être bien informé des manières dont les cookies peuvent être employés.

Il reste aussi le problème des cookies déjà présents sur le disque dur (ainsi que ceux que l'internaute laisse volontairement passer de temps en temps). Même après avoir désactivé l'enregistrement des cookies dans le navigateur, celui-ci enverra malgré tout les cookies existant préalablement à leur interdiction de séjour. Dans ce cas, il n'y a pas d'autre choix que de les effacer manuellement (ou grâce à un logiciel comme certains de ceux évoqués ci-dessus) ou de les modifier et les enregistrer ensuite en read-only pour les rendre inutilisables par les sites Web qui les ont déposés (mais cela n'empêchera pas ces mêmes sites d'en déposer d'autres qu'il faudra à nouveau effacer ou modifier !).

Concernant les cookies provenant des « Web bugs » et des bannières publicitaires, les navigateurs de dernière génération permettent une gestion plus fine des cookies en ajoutant une fonction qui permet à l'internaute de rejeter les cookies ne provenant pas du site Web qu'il est en train de visiter, bloquant ceux de serveurs tiers et en autorisant la lecture d'un cookie uniquement par le serveur qui l'a émis (ce qui évite la lecture de cookies de la même société mais qui auraient été émis par différents serveurs des différents sites de cette même société).

II.4.2. Quelques logiciels de gestion des données à caractère personnel

Quand on visite un site Web, on demande en fait au serveur que le contenu du site qu'il héberge soit téléchargé sur notre ordinateur. Or ce site contient une variété de fichiers qui seront transmis à notre machine, ceux-ci incluent les pages HTML proprement dites mais aussi, comme nous l'avons vu, des fichiers que nous ne désirons pas recevoir ou que l'on ne s'attend pas à recevoir (comme des cookies et/ou des annonces publicitaires). Certains sites demandent également à notre ordinateur de délivrer des informations en retour, le plus fréquemment dans le but d'activer des options de personnalisation que nous avons utilisées précédemment. En utilisant les outils suivants, on peut contrôler et restreindre les excès dans l'échange de fichiers entre son ordinateur et les sites que l'on visite. Le plus souvent, le logiciel nous permettra de personnaliser les mesures de protection d'un site à l'autre, permettant des échanges plus importants avec des sites de confiance.

Internet Guard Dog (Shareware, $39.95), publié par McAfee (http://www.mcafee.com) est un programme qui surveille différents paramètres en rapport avec la vie privée sur l'ordinateur de son utilisateur. Il lui fournit diverses manières de contrôler les cookies, les bannières publicitaires, l'activité sur les forums ou les chat et plus encore. Il protège données et vie privée face aux éventuelles intrusions et attaques tentées via Internet. Ce logiciel offre en effet toute une panoplie de contrôles et de mesures de sécurité susceptibles de prévenir des activités illicites. Il donne un contrôle total sur les cookies et les informations personnelles (mot de passe, contacts et données bancaires) transmises discrètement aux serveurs sur Internet. Il permet d'établir des listes de serveurs dont les

cookies sont acceptés ou refusés. Il peut avertir de tout accès non autorisé à Internet et bloquer les bannières publicitaires des sociétés de cybermarketing. Si l'utilisateur le souhaite, Guard Dog efface automatiquement le contenu du cache Internet et des fichiers d'historique à chaque fermeture du navigateur. Construire des listes de fichiers protégés et de programmes autorisés à y accéder fait aussi partie de ses possibilités. Toute tentative d'accès non autorisée résulte en un message d'avertissement qui donne l'opportunité d'accepter ou de refuser l'opération. Protection contre les virus, protection des mots de passe, support spécial de toute activité ActiveX suspecte sont autant de fonctions disponibles. Guard Dog bénéficie par ailleurs d'une documentation très utile à qui souhaite l'utiliser. Il apporte aussi une protection antivirus complète qui scanne automatiquement chaque fichier téléchargé ou chaque pièce attachée à un e-mail avant qu'il ne soit enregistré sur l'ordinateur et élimine les fichiers résiduels (la mémoire cache du browser et le fichier de l'historique des URL visitées) sur l'ordinateur après chaque session sur Internet. Chaque fois qu'un site Web veut placer un cookie sur la machine d'un utilisateur, Guard Dog « aboie » pour l'alerter. A ce moment-là, l'utilisateur choisit d'accepter ou rejeter ce cookie. Mais le programme peut aussi être paramétré pour chaque site visité – on peut bloquer les cookies de façon permanente pour certains sites et permettre de façon permanente ceux d'autres sites. Le logiciel inclut également un manager de mots de passe, qui est capable de retenir les informations de login pour des sites Web afin que l'utilisateur n'ait pas à rentrer ces informations à chaque visite.
Guard Dog avertit aussi l'utilisateur quand ses informations personnelles sont demandées par (ou transmises à) un site tiers.

AdSubtract PRO (Shareware, $29.95, essai gratuit pendant 30 jours) est un produit développé par AdSubtract.com (http://www.adsubtract.com). Il est capable de bloquer les bannières de publicité qui apparaissent dans la fenêtre du browser et qui accompagnent de nombreux sites sur Internet. Ce logiciel peut également contrôler les cookies ainsi que d'autres informations contenues dans les pages Web (sons, images animées,...) et il est capable d'empêcher la transmission de l'adresse de la page référente. AdSubtract se charge dans la barre des tâches du système d'exploitation Windows. Il est d'origine pré-configuré pour bloquer les annonces publicitaires en ligne (quelles soient animées ou non) et les cookies. De plus, il est entièrement paramétrable selon le site Web visité. Par exemple, il y a moyen d'accepter les cookies, de mémoriser les mots de passe ou de recevoir des publicités de sites spécifiques jugés dignes de confiance et vice-versa. AdSubtract bloque toutes les bannières envahissantes avant qu'elles ne soient affichées sur le browser, ce qui apporte, avec le blocage des cookies, une protection efficace de la vie privée et une navigation sur le Web plus rapide et agréable.
Il est également possible de demander à AdSubtract des informations statistiques sur les bannières et les cookies qu'il a bloqués selon nos préférences (par exemple, on peut connaître le nombre de cookies que le logiciel a stoppés pour une page donnée). En outre, on peut demander au programme de nous notifier par une alerte sonore chaque moment où une bannière ou un cookie est intercepté. AdSubstract fournit encore d'autres options de filtrage comme le blocage des bannières de publicité sous forme de fenêtres pop-up, l'arrêt de l'exécution des JavaScripts et des applets Java, mais aussi le nettoyage des fichiers temporaires du navigateur (mémoire cache, historiques,...).
Un produit spécifique pour le blocage des annonces publicitaires est ***AdWiper*** (Shareware, $17.50, http://adwiper.com/) : il filtre le code HTML des pages Web, quand elles sont affichées par le navigateur, pour supprimer les bannières. Pour ce faire, il utilise plusieurs méthodes pour les identifier et les bloquer, par exemple en se basant sur des mots spéciaux dans les liens ou les tags HTML ou en examinant la taille des images. Il informe également

l'utilisateur du nombre de bannières bloquées. En fait, il bloque en moyenne de 70 à 80% des annonces publicitaires.

IDcide Privacy Companion (logiciel gratuit) a été mis au point par IDcide (http://www.idcide.com). Ce « Privacy Companion » (le « compagnon de la vie privée ») est un plug-in pour les browsers qui informe son utilisateur à propos des profilages éventuels réalisés par les sites Web et les annonceurs, et lui permet d'ajuster sa réponse aux diverses activités de profilage qui peuvent survenir. En outre, il permet aussi de savoir quels sites utilisent des cookies et comment ; on peut décider de bloquer, quand on le souhaite, les systèmes de profilage basés sur les cookies. Enfin, il permet à son utilisateur de se prémunir de la « trahison » de son browser vis-à-vis des sites tiers au moyen de la transmission de la page référente, qui indique par exemple les mots-clés recherchés sur un moteur de recherche. Une fois que le plug-in est installé, une série de petites icônes font leur apparition dans le coin supérieur gauche du browser. Ces icônes s'actualisent continuellement afin d'avertir l'utilisateur du niveau du profilage effectué à la fois par le site qu'il visite mais aussi par des tierces parties (comme les annonceurs) présentes sur le site. A chaque instant, l'utilisateur a l'opportunité de choisir entre trois niveaux de protection de sa vie privée – aucune protection (niveau bas), protection contre le profilage par des tierces parties (niveau moyen), et protection contre le profilage exercé par le site visité et par des tierces parties (niveau élevé). Le niveau moyen permet à l'utilisateur de bénéficier des avantages de personnalisation du site Web visionné tout en se protégeant simultanément de la surveillance des annonceurs. Le logiciel donne aussi des informations statistiques à propos des sites qui tentent d'espionner l'internaute. Ces informations comportent les noms des sites Web qui effectuent un profilage, avec qui ils sont associés pour pister et recueillir des informations à caractère personnel, et le nombre de fois qu'ils ont demandé des informations à propos de l'utilisateur ou de son navigateur. Un prototype récent de ce logiciel permet même à l'utilisateur de mettre au point un système de décisions automatiques basé sur les politiques P3P publiées par certains sites.

II.4.3. Diverses technologies de protection de la vie privée

a) l'OPT-OUT ou la suppression du numéro d'identification contenu dans les cookies d'une société de cybermarketing

L'opt-out est un terme qui illustre une démarche active : « je ne veux pas recevoir d'informations de votre part ». Difficile d'envoyer des e-mails à tous les sites qui ont recours aux cookies. Il faut donc se contenter d'avertir quelques-uns des plus gros utilisateurs de cookies : les régies publicitaires. La Network Advertising Initiative (http://www.networkadvertising.org) est une association américaine regroupant 7 des plus grosses régies d'Internet. Elle propose aux internautes de s'inscrire sur une « liste rouge » de cookies. Cela leur évitera d'accumuler des cookies dont les émetteurs ne sont pas directement les sites qu'ils visitent.

Les sites Web et régies publicitaires qui placent un cookie sur le disque dur des internautes y inscrivent souvent un numéro unique (le Global Unique Identifier), qui permet d'identifier leur ordinateur chaque fois que ce cookie leur est retransmis. Néanmoins, certaines firmes de cybermarketing permettent de mettre à ce numéro une valeur spéciale indiquant que l'internaute en question ne souhaite plus être profilé par cette firme. C'est le cas de Doubleclick, l'une des régies de publicité en ligne les plus prisées des sites Web. Pour accéder à la procédure dite d'Opt-Out, il faut se rendre à l'adresse suivante : http://www.doubleclick.com/us/corporate/privacy/privacy/default.asp et cliquer sur le bouton

« Ad Cookie Opt-out ». Le numéro d'identification contenu dans le cookie est remplacé par la mention Opt-Out. http://www.angelfire.com/va2/webdesignclass/optoutcookie.htm est un site permettant de mettre en œuvre la procédure d'Opt-Out pour d'autres firmes de cybermarketing.

b) Modification du fichier "*Hosts*"

Le fichier « Hosts » dans Windows[90] (mais aussi dans d'autres systèmes d'exploitation, comme Linux ou MacOS) est utilisé pour associer les noms de domaine avec les adresses IP correspondantes. Rappelons qu'en réalité, les ordinateurs n'utilisent que les adresses IP pour trouver les sites, tandis que nous, nous utilisons les noms de domaine[91], afin de ne pas avoir à nous souvenir de longues chaînes de chiffres quand nous voulons visiter un site (les mots étant plus facilement mémorisables). Quand un internaute demande l'affichage d'une page Web, son navigateur doit, avant d'accéder au site correspondant, traduire les noms de domaine en adresse IP (ce ne serait pas nécessaire si nous utilisions directement des adresses IP). Or une série d'étapes est nécessaire pour traduire un nom de domaine en son équivalent sous la forme d'une adresse IP. La première, et la seule qui nous concerne ici, est le fichier *Hosts* sur l'ordinateur local. Ce fichier réalise l'association nom de domaine/adresse IP pour un certain nombre, limité, de noms de domaine distincts. Si l'adresse IP recherchée est trouvée dans le fichier *Hosts* (ce qui veut dire que le nom de domaine indiqué y était mentionné), le navigateur peut aller directement sur le site en question. Sinon, le navigateur devra interroger un serveur DNS pour obtenir l'information. Puisque la recherche se termine une fois que la correspondance est trouvée, cela fournit un mécanisme pour bloquer les sites jugés suspects (ou plutôt empêcher une communication avec les serveurs les hébergeant). Par exemple, on pourra bloquer les sites qui servent les annonces publicitaires, les sites dont le contenu est condamnable, ou n'importe quel autre site que l'on estime utile de bloquer.

En fait, rien ne nous empêche de mettre de nouveaux noms de domaine et les adresses IP correspondantes dans le fichier *Hosts* en le modifiant avec un simple éditeur de texte. De plus, il existe sur Internet de telles tables de correspondance toutes prêtes qui peuvent être téléchargées[92] par toute personne désireuse d'enrichir la leur (attention toutefois de sauvegarder préalablement son ficher existant avant de l'écraser par un nouveau, afin de pouvoir revenir en arrière si nécessaire...). Ce faisant l'ordinateur n'ira plus interroger un serveur DNS pour traduire ces noms de domaine. Ceci améliorera la vitesse d'accès aux sites en question puisque l'ordinateur gagne le temps nécessaire au questionnement d'une autre machine située sur Internet pour effectuer la traduction.

Tous les ordinateurs ont un nom de domaine qui leur est propre – il est appelé « localhost » et est associé à l'adresse IP locale *127.0.0.1* – et qui est utilisé pour référer à lui-même. Or si on associe, au moyen du fichier *Hosts*, un autre nom de domaine avec l'adresse IP locale, on a effectivement bloqué ce nom de domaine et donc, le site correspondant, puisque toutes les tentatives d'accès à ce site conduiront inévitablement à l'ordinateur local. C'est de cette façon que l'on bloque des sites en utilisant le fichier *Hosts* : en disant à notre

[90] Ce fichier se trouve dans le répertoire *c:\windows*, pour les systèmes Windows 9x/ME, ou dans *c:\winnt\system32\drivers\etc.*, sous Windows NT/2000/XP.

[91] Par exemple, le nom de domaine de *Yahoo!* est www.yahoo.com, et son adresse IP est *204.71.200.67*

[92] Sur http://www.accs-net.com/hosts par exemple, on peut trouver un fichier contenant les adresses de plus de 9400 serveurs jugés douteux par son auteur.

ordinateur que l'adresse IP du site que nous voulons bloquer est la sienne. On a donc court-circuité la traduction du nom de domaine pour indiquer l'adresse IP que l'on souhaite. L'ordinateur ne pourra jamais accéder au site en question puisqu'il ne l'héberge pas, alors qu'il pense pourtant l'avoir effectivement trouvé. Un message d'erreur s'affichera alors dans la fenêtre du navigateur indiquant que le serveur hébergeant le site ne répond pas.

Beaucoup de sites Web sont liés à des serveurs tiers pour l'obtention des bannières publicitaires. Si le fichier *Hosts* a été modifié comme indiqué pour ces serveurs, le navigateur échouera rapidement lors de sa recherche des données requises (à savoir les bannières) à partir des serveurs de publicités (il ne les trouvera jamais puisqu'il cherchera au mauvais endroit tout en étant persuadé du contraire), et continuera à charger les portions pertinentes de la page visitée. Cette méthode nous permet d'empêcher notre ordinateur d'entrer en communication avec les serveurs des annonceurs, et par conséquent, nous ne verrons plus leurs bannières. De plus, les annonceurs ne pourront plus nous envoyer des cookies et par là, établir notre profil.
Il existe certains logiciels, comme ***DNSKong*** (http://www.pacificnet.net/~bbruce/system.htm, gratuit pour un usage non commercial), qui permettent de faciliter la gestion du fichier *Hosts*.

Bénéfice de cette méthode :

1. Elle utilise très peu de ressources.

En utilisant une fonction incorporée dans notre système d'exploitation et utilisée par lui de façon systématique, nous sommes capables de bloquer les sites des annonceurs (ou de n'importe quel autre site si nous le désirons) sans avoir besoin d'un logiciel supplémentaire. Cela réduira l'utilisation du processeur et de la mémoire vive, qui seront libres pour d'autres tâches.

2. Elle opère également pour d'autres protocoles que HTTP.

La plupart des programmes bloquant les bannières publicitaires interceptent uniquement les communications IP allant vers le port HTTP. Le fichier *Hosts*, cependant, bloquera les communications IP quel que soit le port (et donc, quel que soit le protocole applicatif utilisé).

3. Elle permet d'empêcher l'utilisation de plusieurs techniques de profilage.

En interceptant les communications IP avant qu'elles ne quittent notre ordinateur, le fichier *Hosts* peut empêcher les compagnies de publicité en ligne (ou, plus généralement, celles qui effectuent un pistage des internautes) de nous « connaître » lorsque nous naviguons sur le Web. Cela nous prémunira contre leur profilage et nous aidera à protéger notre vie privée. Tous les sites dans le fichier *Hosts* associés à l'adresse *127.0.0.1* ne seront jamais atteints. Par contre, les sites qui ne sont pas dans ce fichier pourront continuer à nous tracer et à nous envoyer des annonces publicitaires.

4. Le fichier *Hosts* est configurable.

Au lieu de se fier aux jugements d'autres personnes pour décider des sites qui méritent ou pas d'être bloqués, on peut éditer le ficher *Hosts* entièrement par soi-même. Cela signifie que nous pouvons mettre dans ce fichier uniquement les sites que nous désirons et qu'ils ne seront plus accessibles ensuite à partir de notre ordinateur (on peut de la sorte protéger ses jeunes enfants, par exemple).

5. La vitesse de navigation est accrue.

Comme nous l'avons déjà évoqués ci-dessus, en plaçant des sites dans notre fichier *Hosts* avec leurs adresses IP correctes, notre ordinateur n'a plus besoin de demander à une autre machine où se trouve le site recherché et y accède directement. Ceci peut accroître significativement la rapidité de la navigation sur les sites que l'on fréquente le plus régulièrement. En outre, en empêchant les publicités en ligne d'être téléchargées (par le blocage des serveurs des annonceurs), les pages Web sont visibles plus rapidement.

Quelques restrictions à cette méthode :

1. Elle n'opère pas pour des noms de domaine, tels que *.cequevousvoulez.com (ce que l'on appelle les « wildcards »).

2. Elle ne fonctionne pas avec des URLs qui commencent par des adresses IP.

En effet, cette méthode ne s'applique qu'aux noms de domaine, pas aux adresses IP. La raison est que le fichier *Hosts* est utilisé pour déterminer une adresse IP à partir d'un nom de domaine, et non l'inverse. Donc, si le navigateur connaît déjà l'adresse IP (parce qu'elle est communiquée dans l'URL), le fichier *Hosts* n'est pas consulté et ainsi, ne peut pas bloquer le site en question (le navigateur s'y rendant directement).

3. Elle n'agit pas sur les publicités qui sont servies par le même serveur que le site visité.

En effet, le fichier *Hosts* ne peut bloquer que des sites entiers (en fait, les serveurs hébergeant les sites) et n'est pas capable de bloquer des sous-répertoires ou des fichiers particuliers sur un site. Par exemple, il est impossible de bloquer *www.doubleclick.com/ads* parce qu'il s'agit d'un sous-répertoire. Il faudrait alors bloquer le nom de domaine *www.doubleclick.com*, ce qui rendrait impossible la navigation sur l'entièreté du site (les bannières étant, dans ce cas de figure, des fichiers contenus dans un sous-répertoire du site tout comme les pages HTML). Par conséquent, il faut utiliser autre chose pour éliminer les annonces issues du site consulté, comme par exemple, un programme spécifique de blocage de publicités.

4. Elle peut causer un dysfonctionnement de certains sites.

En effet, si on place dans le fichier *Hosts* une correspondance erronée (le nom de domaine d'un site étant associé à l'adresse IP d'un autre site), cela impliquera que le site en question sera « remplacé » par un autre et donc ne sera plus visible normalement. Pour remédier à cela, il faut soit corriger la mauvaise correspondance, soit l'effacer du fichier. Il est d'ailleurs préférable de ne garder dans le fichier *Hosts* que les noms de domaine des sites que l'on visite fréquemment ou de sites que l'on désire bloquer (et de les supprimer du fichier lorsque l'on souhaite débloquer les sites correspondants).

En conclusion, en créant lui-même sa propre table de correspondances, l'internaute peut éviter ainsi les connexions à un serveur DNS pour ses sites favoris, ce qui lui permet de gagner du temps ; d'autre part, il peut aussi bannir l'accès à certains sites. L'astuce consistant à indiquer, en face du nom du site à bloquer, l'adresse IP de son propre ordinateur (*127.0.0.1*, l'adresse « localhost »). Dès que le navigateur lit cette adresse, il est incapable de s'y connecter.

c) Effectuer confidentiellement des recherches sur Internet

Parmi les outils les plus importants et les plus populaires sur le Web, on trouve les moteurs de recherche. Ces sites font qu'il est désormais aisé de trouver rapidement l'information que l'on désire consulter sur Internet. Afin de rendre leur service encore plus utile et personnel, beaucoup de moteurs de recherche utilisent à présent des cookies et d'autres moyens de collecte de données à caractère personnel comme une partie intégrante de leur infrastructure.
Les internautes qui désirent utiliser un moteur de recherche de haute qualité tout en maintenant simultanément un haut niveau de protection de leur vie privée en ligne peuvent maintenant se servir du produit suivant.

TopClick Private Web Search a été créé par la société TopClick International (http://www.topclick.com) et est consultable gratuitement. C'est un moteur de recherche sur Internet, similaire dans son efficacité à Google (puisque TopClick utilise la technologie de recherche mise au point par Google), à la différence que TopClick garantit qu'il ne distribuera jamais une bannière publicitaire non sollicitée aux internautes l'employant et qu'il ne fera pas l'usage des cookies ou d'autres technologies qui ont la capacité de créer un profil non désiré des utilisateurs ou de révéler des informations personnelles à son sujet à des tiers. La politique de TopClick en matière de vie privée a été vérifiée par des organisations tierces comprenant BBB Online[93]. En cas de changement de la politique du site, les utilisateurs pourront avoir le choix de décider comment et quand les informations personnelles qu'ils fournissent à TopClick seront utilisées. En d'autres mots, TopClick donnera à ses utilisateurs l'option d'opt-in pour la collecte et la distribution des données à caractère personnel.
TopClick dispose également d'un « Privacy Resource Center », qu'un internaute peut utiliser pour obtenir des conseils ou des nouvelles à propos de la protection de la vie privée sur Internet. TopClick vend du matériel dans le domaine de la confidentialité, ce qui a pour but de financer leur site pour abandonner totalement l'usage de la publicité. Le site a également l'intention de vendre un logiciel qui donnera à ses futurs utilisateurs les moyens d'administrer leurs courriers électroniques et d'éliminer le Spam.

e) Navigateur avec technologies de protection de la vie privée intégrées

FlowProtector (http://www.flowprotector.com/) est un navigateur sécurisé en français qui permet de surfer sans laisser de traces. Il offre des fonctions de firewall personnel, de gestion sélective des cookies par site, de blocage des bannières publicitaires et des Web bugs, de gestion et protection des informations personnelles et de contrôle parental (système de liste noire et liste blanche). De plus, il ne transmet pas les URLs des sites consultés à d'autres applications et inclut un système de cotation des sites marchands. Enfin, il est doté d'un système de protection contre les virus trojans, les contenus actifs (contrôles ActiveX et applets Javas) dangereux et les codes malicieux.

[93] Better Business Bureau Online (http://www.bbbonline.org/) est une organisation sans but lucratif dont la tâche est d'aider les consommateurs et les entreprises à se maintenir dans un marché en ligne éthique et à déclarer des pratiques respectueuses de la vie privée.

f) Bugnosis ou la détection des Web bugs au cours de sa navigation

La Privacy Foundation, une association américaine militant activement pour la protection des données personnelles, a mis au point un petit programme gratuit baptisé ***Bugnosis*** (http://www.bugnosis.org/) qui détecte les Web bugs, ces mouchards invisibles présents sur certains sites.
La fenêtre de Bugnosis se greffe sur Internet Explorer. Il lui ajoute une fenêtre affichant pour chaque page visitée le compte rendu complet de ce qui y est dissimulé. Une alerte sonore peut être paramétrée pour signaler la page Web suspecte, d'autres options sont également disponibles, comme l'affichage du contenu des cookies tiers. Bugnosis « concrétise » alors dans le corps de la page l'endroit exact où se dissimule un fichier invisible à l'écran. Sa position est alors désignée par une icône clignotante à l'effigie d'un insecte.

Le logiciel détecte principalement les Web bugs, ces petits espions qui sont dissimulés dans une page Web sous la forme d'une image d'un pixel de côté (donc invisible à l'écran), et dont la vocation est généralement de compter les visites sur une page. Mais il détectera également toutes les images d'une page Web qu'il considère comme des Web bugs, donnant pour chacune d'entre elles une description détaillée de ce qui a éveillé ses soupçons : image trop petite, image renvoyant à un autre site que celui qui est visité, image associée à un cookie, etc. Cependant, comme pour certains programmes de détection, les alarmes sont nombreuses et pas toujours pertinentes... La Privacy Foundation a semble-t-il ratissé large. Il faut donc faire le tri, ce qui exige un minimum de connaissances techniques. Mais l'utilisation de Bugnosis réserve parfois de réelles surprises.

II.4.4. Analyser les politiques en ligne en matière de vie privée

Les politiques en matière de vie privée sont des outils utiles pour les consommateurs. En examinant la déclaration d'un site en matière de vie privée, on peut déterminer si les politiques de collecte de données de l'organisation concordent avec nos préférences dans ce domaine. Bien que l'examen de la politique d'un site puisse se révéler utile, cela prend aussi du temps pour accéder à l'information de chaque page que l'on visite.
Les produits suivants simplifient le processus d'inspection des politiques des sites visités en distillant les principes d'un site en matière de vie privée en un rapport « photographique ». Par un certain nombre de méthodes, on peut déterminer rapidement si le niveau de protection d'un site est bon pour soi.

La ***Platform for Privacy Preferences (P3P)*** (http://www.w3.org/P3P/), en français « Plate-forme de préférences en matière de protection de la vie privée », a été créée par le World Wide Web Consortium (W3C). Cet organisme, composé de plus de 400 leaders de l'industrie informatique, se consacre à la formulation des standards technologiques sur Internet qui s'adressent aux questions de politique publique, sociale et légale. Le consortium a élaboré un nouveau standard qui a pour but de simplifier grandement les contrôles individuels des pratiques en matière de vie privée des sites Web, ce standard se nomme P3P et il est basé sur la technologie XML. L'objectif de P3P est de permettre aux sites Web d'exprimer leurs préférences en matière de protection de la vie privée, et aux utilisateurs de faire valoir leurs préférences à ce sujet, de telle sorte que les utilisateurs soient à même de prendre en toute connaissance de cause des décisions quant à la manière dont ils utilisent le Web et de contrôler l'usage fait de leurs informations.
P3P offre avant tout aux webmasters une manière simple et aisée pour publier la politique en matière de vie privée de leur site dans un format standard compréhensible pour les

ordinateurs. De plus, P3P peut être déployé sur un site Web existant sans nécessiter de profondes modifications. Cette technologie devrait aboutir au développement d'outils, intégrés aux navigateurs ou dans des logiciels séparés, permettant de prendre une « photographie » de la politique d'un site, de la comparer avec les préférences de l'internaute en matière de vie privée, d'avertir celui-ci sur le résultat de cette comparaison et éventuellement le conseiller sur la marche à suivre dans le cas où il y aurait désaccord. Ainsi, ils simplifieront les négociations en matière de vie privée entre les internautes et les sites Web. De plus, les clients P3P seraient capables de contrôler la politique d'un site chaque fois qu'elle est modifiée. De plus, ils pourraient effectuer ce contrôle sur chaque élément d'une page, y compris les bannières publicitaires issues de sites tiers ou les images invisibles. Avec P3P, les sites Web peuvent imbriquer leurs politiques en matière de vie privée dans chacune de leurs pages, en les ajustant à l'intérieur des données de chaque page. Par exemple, la page d'accueil d'un site commerçant ne peut pas demander les mêmes informations personnelles que ses pages d'achat, et la programmation de chaque page reflètera cette différence. Les internautes disposant d'un browser en mesure d'interpréter le langage P3P peuvent y indiquer leurs préférences en matière de vie privée avant de surfer sur Internet (pour ce faire, ils remplissent un questionnaire standardisé indiquant sous quelles conditions leurs informations personnelles peuvent être collectées, sauvegardées et éventuellement données ou vendues à des tiers). Le navigateur analysera alors le code P3P imbriqué dans une page et comparera le résultat avec les préférences définies par l'internaute. Si les deux concordent, la navigation de l'internaute continuera sans interruption. Par contre, si la politique de la page est en contradiction avec les choix de l'internaute, le navigateur l'avertira et lui donnera la possibilité de ne pas continuer plus loin sa navigation.

Voyons comment le standard P3P est sensé améliorer la protection des données à caractère personnel. Avec l'arrivée des navigateurs de dernière génération, il est maintenant possible de contrôler l'utilisation de ses données personnelles grâce à la technologie P3P.
La plateforme P3P se compose de deux parties :

1) La première partie se situe au niveau du navigateur utilisé par l'internaute. Dans les préférences, il est possible de définir ce qu'on appelle la politique d'utilisation des données personnelles, càd quelles sont les données personnelles que l'utilisateur lui confie ainsi que les finalités pour lesquelles ces données peuvent être communiquées à des sites Internet (ce sont les *privacy preferences* de l'utilisateur).
2) La deuxième partie de la plateforme P3P, réside au niveau du site visité. En effet, l'administrateur du site doit déclarer une politique P3P qui définit les finalités pour lesquelles il souhaite utiliser les différents types de données et comment celles-ci sont utilisées (ce sont les *privacy practices* du site). Lorsqu'un internaute visite le site, son navigateur interroge (de façon invisible pour l'internaute) la politique du site pour savoir si elle est en corrélation avec celle définie dans ses préférences. S'il y a accord entre les deux, les données correspondantes seront transférées ; sinon (càd si le site n'a pas de politique P3P ou si sa politique est plus faible que les préférences de l'utilisateur), elles ne le seront pas, une alerte sera affichée et il pourra y avoir une négociation sauf si l'internaute désire quitter le site immédiatement. S'il désire continuer sa navigation sur le site, il saura que ce site ne satisfait pas ses préférences.

Ainsi, l'utilisation des cookies devient conditionnée par la politique P3P définie par l'internaute. Il est alors nécessaire pour les sites utilisant des cookies de déclarer une politique P3P. Sinon, par défaut, l'utilisation des cookies sera interdite par le navigateur de l'internaute.

Quelles conditions devraient être remplies par toute plate-forme technique de protection de la vie privée sur le Web, dans le but d'éviter la récolte systématique de données à caractère personnel ?

1. La technologie ne peut être une solution en soi à la protection de la vie privée sur le Web. Elle doit être appliquée dans un cadre réglementaire.
2. Tout utilisateur doit pouvoir surfer anonymement sur le Web, s'il le désire.
3. Avant que des données à caractère personnel, et en particulier celles dévoilées par l'utilisateur, ne soient traitées par le fournisseur d'un site Web, l'accord informé de l'utilisateur est nécessaire. De plus, certaines règles de base et auxquelles il ne sera jamais dérogé, devraient être incluses dans la configuration par défaut de la plateforme technique.

Cette plate-forme doit être mise en œuvre dans le contexte d'un cadre de règles de protection des données qui fournissent un niveau minimal et non négociable de protection de la vie privée à tous les individus. Pour cela, il reste encore un nombre important de questions à résoudre. C'est lorsque ces questions auront été résolues que P3P pourra jouer un rôle positif, s'il est mis en œuvre dans un cadre adéquat.
Signalons tout de même que les principaux avantages de P3P sont les suivants :

- P3P peut aider à normaliser les notices concernant la protection de la vie privée. Bien qu'en soi, ceci n'apporte pas de protection de la vie privée, son instauration pourrait sensiblement promouvoir la transparence, et soutenir les efforts de renforcement de la protection de la vie privée.
- P3P peut soutenir le développement des choix en matière de protection de la vie privée, y inclus l'anonymat et l'usage de pseudonymes.

Il faut cependant tenir compte des limitations de P3P, qui sont:

- P3P ne peut pas assurer la protection de la vie privée des utilisateurs dans des juridictions ne disposant pas de lois suffisantes sur la protection de la vie privée; il n'est pas à même de créer des politiques publiques et ne peut pas non plus exiger que ses spécifications soient suivies sur le marché.
- P3P ne peut pas garantir que les sociétés respecteront des politiques de protection de la vie privée, P3P supposant implicitement que la solution à la protection des données à caractère personnel est la déclaration des pratiques de ces données par le site. Cela suppose une honnêteté absolue de la part de tous les sites dans un contexte où aucune sanction n'est envisagée (des sanctions au non-respect des promesses ne peuvant être prévues que par la loi ou par l'adhésion à un organisme qui assure sa propre police). En fait, P3P est incapable de garantir que le site fait en réalité ce qu'il prétend faire.
- P3P n'est pas une technologie suffisante en soi pour assurer la protection de la vie privée sur Internet, elle ne propose pas une solution globale aux autres problèmes de protection des données sur le Web.

En fait, P3P dans son stade actuel n'offre qu'une solution partielle reposant sur un code de « bonne » conduite, invisible de l'internaute. De plus, selon cette plate-forme, c'est l'internaute lui-même qui doit spécifier et déclarer ses préférences en matière de vie privée (ce qui implique qu'il soit suffisamment informé sur la question ; de surcroît, ces préférences sont elles-mêmes des données personnelles et il faut aussi s'assurer que des personnes soucieuses de préserver leur vie privée ne se retrouvent fichées pour cette raison, or rien ne le garantit !), alors que cela devrait être le responsable du traitement des données à caractère personnel qui devrait renseigner les internautes de ses pratiques et de leurs finalités.

II.4.5. Les Infomédiaires

Une personne peut également décider de passer par les services d'un Infomédiaire. Les Infomédiaires peuvent être décrits comme suit : « Un Infomédiaire, ou intermédiaire d'information, est une personne de confiance ou une organisation ayant accès au Web, qui est spécialisée dans des services d'information et de connaissance destinés à/au sujet de/pour le compte d'une communauté virtuelle. L'Infomédiaire facilite et stimule la communication et l'interaction intelligentes parmi les membres d'une communauté virtuelle. Elle administre et développe un capital de connaissance privée, qui inclut des contenus et des hyperliens, qui sont d'un intérêt particulier pour la communauté ».
Conformément aux contraintes de respect de la vie privée qui sont exigées par la communauté virtuelle, l'Infomédiaire réunit, organise et publie de façon sélective des informations sur la communauté et ses membres dans le but de satisfaire les besoins de la communauté virtuelle...

L'Infomédiaire est un nouveau genre d'intermédiaire d'affaires qui aide ses clients à capturer, gérer et maximiser la valeur de leurs données à caractère personnel. Les consommateurs ont montré qu'ils étaient prêts à dévoiler des données à caractère personnel s'ils pouvaient en tirer un avantage, mais ils se rendent de plus en plus compte qu'ils vendent leur vie privée à trop bon marché, à des sociétés qui les utilisent pour réaliser du profit à leur propre avantage. En un mot, ils trouvent insatisfaisants les avantages retirés de la divulgation de leurs informations personnelles. Les Infomédiaires peuvent aider les consommateurs à faire de meilleures affaires avec les sociétés, en agrégeant leurs données avec celles d'autres consommateurs, et ainsi tirer parti de leur position de force en négociant pour leur compte avec les sociétés. Ils agissent en tant que gardiens, agents et courtiers d'informations concernant des consommateurs, les vendent aux sociétés (en leur y donnant accès) pour le compte de ces consommateurs, tout en protégeant leurs données à caractère personnel contre tout abus.
L'aspect positif de l'Infomédiaire est que, dans de nombreux cas, il peut acheter les biens ou services requis et les livrer au consommateur final, tout en lui permettant de conserver son anonymat. La société Infomédiaire peut également fournir des agents intelligents pour aider les souscripteurs à réaliser leur tâche.

Les clients des Infomédiaires peuvent théoriquement rester éternellement anonymes, pendant qu'ils surfent sur le Web ou font des achats en ligne. Ils seront cependant sollicités à n'en rien faire, parce qu'ils toucheront des vendeurs une légère redevance chaque fois qu'ils consentent à divulguer leur identité ou leur adresse e-mail. Cette redevance peut être un paiement en argent ou une réduction sur le prix du produit acheté.
Les clients recevront aussi de l'argent chaque fois qu'ils fourniront à certaines sociétés un accès à leur profil personnel. Le montant concerné dépendra des préférences de chaque client en matière de respect de sa vie privée. Les clients qui choisissent de rester totalement anonymes renoncent à recevoir de l'argent, en échange de l'assurance de la protection de leur vie privée. Les clients qui font confiance aux contrôles imposés par l'Infomédiaire sur l'accès à leurs informations, et qui comprennent l'intérêt d'une divulgation sélective à des sociétés, peuvent gagner de l'argent.

Des sociétés comme Lumeria (http://www.lumeria.com/) ou Privista (https://www.privista.com/) réunissent les principales caractéristiques de l'infomédiaire. Privista est une société apparue en septembre 2000. Elle offre aux consommateurs un service gratuit pour retirer leurs coordonnées des grands fichiers du marketing direct. Lumeria propose le même type de services que Privista et va même jusqu'à menacer d'engager des

actions en justice si les détenteurs de fichiers se font un peu tirer l'oreille pour obtempérer. Pour gagner de l'argent, Lumeria espère demander à l'avenir aux services marketing des entreprises d'acheter les profils détaillés de « leurs » clients mais pas leur identité. A terme, des actions marketing pourront ainsi être menées auprès de cibles extrêmement précises sans que les émetteurs des messages et des propositions n'aient à connaître l'identité des destinataires. Ces sociétés contribueront peut-être à protéger réellement la vie privée de leur client mais force est de constater qu'il s'agit plutôt pour elles d'un moyen et non d'une fin. Leur objectif est bien de redessiner entièrement à leur profit le paysage de l'activité des bases marketing. Pour ce faire, elles amassent des données personnelles, vendent des profils et ne divulguent pas l'identité réelle de leurs abonnés.

En conclusion, on peut dire que si les Infomédiaires peuvent jouer un rôle positif dans la protection des données à caractère personnel des utilisateurs avec lesquels ils ont une relation de confiance, la base d'un tel accord est la possibilité de gagner de l'argent en divulguant ou en donnant accès aux données à caractère personnel des consommateurs.
Selon les circonstances et la nature de l'Infomédiaire, le système peut donc avoir un caractère augmentant, ou diminuant, la protection de la vie privée.

II.4.6. Les Labels de protection de la vie privée

Labelliser consiste à appliquer un sceau de qualité à un site Web. Depuis plusieurs années, différents labels de protection de la vie privée sont apparus: TRUSTe[94] (http://www.truste.com/), Privaseek (http://www.privaseek.com), the Better Business Bureau (http://www.bbbonline.org/), WebTrust (http://www.cpawebtrust.org/) sont des exemples de tels labels. Ces organisations américaines ont pour but d'être actifs internationalement et donc aussi en Europe, ce qui est déjà le cas pour certaines d'entre elles. Au même moment, des initiatives similaires sont entreprises en Europe, avec un même objectif international, comme par exemple, L@belsite en France (http://www.labelsite.org/).

Un label de protection de la vie privée est accordé aux sociétés qui remplissent un certain nombre d'exigences spécifiées par l'organisme accordant le label. Un tel organisme peut exercer un certain contrôle sur le respect des politiques de protection de la vie privée publiées par les sociétés détentrices du label, par des contrôles périodiques des activités de ces sociétés. Dans certains cas, l'organisme accordant le label traite également les plaintes déposées par des consommateurs contre des sociétés détentrices du label sur leur site Web.

Ces labels ont pour tâche de faire le compte-rendu des politiques en matière de vie privée de sites commerciaux ou non-commerciaux afin de fournir aux internautes une « mesure » facilement compréhensible de ce qu'un site spécifique fera ou ne fera pas avec leurs données à caractère personnel. Les informations contenues dans ce compte-rendu indiquent si l'internaute peut être sujet au Spam, s'il peut directement contrôler quelles informations seront collectées par le site et si le site Web offre le choix d'opt-in (l'internaute doit donner expressément sa permission au site pour que ce dernier communique ses informations personnelles à des tiers) ou au contraire d'opt-out (l'internaute doit expressément choisir de ne pas permettre au site de partager les données personnelles que ce dernier a collectées avec des tiers).

[94] TRUSTe est une organisation indépendante et sans but lucratif aidant les consommateurs et les sociétés à établir des relations de confiance basées sur le respect de l'identité et des informations personnelles. Pour ce faire, elle a mis au point un label qui garanti une utilisation respectueuse des données et informations personnelles tant aux individus qu'aux organisations.

Le label de protection de la vie privée suscite un certain nombre de questions :

1. La première question est celle du contenu du label.

Les droits d'information, d'accès et de refus sont quelques-unes des pierres angulaires des principes européens relatifs à la protection des données et de la vie privée. Le principal risque pour la communauté des internautes est de voir disséminés en Europe un grand nombre de labels de protection de la vie privée, ce qui pourrait d'une manière ou d'une autre induire en erreur les utilisateurs et les responsables de données. Alors qu'ils en donnent l'impression, tous les labels n'apportent pas une garantie sérieuse de respect des principes de protection des données.

2. Le second problème réside dans le contrôle des pratiques de respect de la vie privée du site Web.

De nombreux contrôles peuvent être envisagés. Mais certaines préoccupations à ce propos peuvent être mentionnées :

- Qui effectuera le contrôle (qui, comment, disposant de quel pouvoir vis-à-vis de la société contrôlée) ? Dans le pire des cas, il semble que le contrôleur sera, en première instance, l'intéressé lui-même, pour lequel il sera difficile d'identifier les différences entre la réalité et les pratiques mentionnées, de les prouver, et d'en faire rapport à la société accordant le label. Qui plus est, tous les organismes accordant des labels ne peuvent garantir que les sociétés font réellement ce que dans leurs politiques elles prétendent faire.
- Qui paiera ? Du fait que les organismes accordant un label sont des initiatives privées et ne bénéficient souvent d'aucun soutien financier du gouvernement, certains de ces organismes seront mis sous pression par les sociétés qu'ils sont censés contrôler.
- S'il y a des sanctions, quelles seront-elles ? Toute transgression appelant une sanction adéquate.

Les effets des labels de protection de la vie privée sur le renforcement de la protection de la vie privée ne doivent cependant pas être sous-estimés, car ils peuvent accroître la prise de conscience du respect de la vie privée chez les internautes.

Pour aborder les problèmes mentionnés ci-dessus, certaines propositions peuvent être faites:

a. Contenu du label : Dans le but de garantir que les labels sont conformes à la législation européenne sur la protection des données, des normes européennes pour ces labels pourraient être conçues. Ces normes spécifieraient les exigences auxquelles doivent répondre les labels. Différents labels pourraient coexister, dans la mesure où les internautes pourraient savoir clairement ceux qui répondent aux normes européennes.

b. Contrôle des pratiques des sites Web : La fiabilité des pratiques des sites Web, en matière de protection de la vie privée, pourrait être sensiblement renforcée en obligeant les sites Web munis du label à se soumettre à des audits périodiques. Les normes européennes concernant les labels pourraient inclure cette obligation d'audit et en définir les modalités pratiques, comme un audit par la société elle-même selon une check-list standard, ou un audit par un tiers, etc.

II.4.7. Les Tunnels SSH, les Proxies Socks et HTTPort

a) Les Tunnels SSH

Nous avons vu que malgré l'utilisation de proxies, l'anonymat est tout relatif car un fournisseur d'accès Internet, bien que ne connaissant pas le détail des navigations de ses clients, peut encore intercepter et analyser les données transitant entre les ordinateurs de ces derniers et le proxy. De plus, de tierces personnes peuvent également accéder et surveiller les données en transit sur le réseau. Si l'on ne peut empêcher cette interception, on peut faire en sorte que les données interceptées ne puissent pas être lues. Il faut donc les chiffrer pour en assurer le secret.

Un tunnel SSH[95] est une liaison chiffrée entre un ordinateur et un service distant qu'il utilise comme proxy. Les données envoyées à un proxy devront être chiffrées systématiquement par l'ordinateur de l'utilisateur, le proxy les déchiffrera et les enverra à leur destinataire. De la même manière, les données qui seront renvoyées vers l'utilisateur seront d'abord chiffrées en temps réel par le proxy par qui elles transiteront, celui-ci les lui envoyant ensuite. Ces données seront donc protégées vis-à-vis du fournisseur d'accès de l'internaute ou pour toute personne voulant intercepter les communications entre l'utilisateur et le proxy, puisque tout « curieux » interceptant les données ne verra que deux choses :

1 - toutes les données sont envoyées à un unique service, ou proviennent d'un seul service (le proxy),

2- toutes les données sont chiffrées, et donc illisibles.

---

[95] SSH (Secure SHell) est à la fois la définition d'un protocole et un ensemble de programmes utilisant ce protocole, destinés à permettre aux utilisateurs d'ouvrir, depuis une machine cliente, des sessions interactives à distance sur des serveurs et de transférer des fichiers entre les deux.
Parmi les avantages de SSH sur des applications analogues (telnet, rlogin, ftp) :

- la connexion entre le client et le serveur est chiffrée (cette connexion utilise généralement le port 22). Les hackers, même s'ils contrôlent cette connexion, ne peuvent pas lire les données confidentielles (telles que des mots de passe) qui y transitent.
- la cryptographie à clefs publiques permet à SSH de garantir l'authentification mutuelle du client (utilisateur) et du serveur. Des hackers ne peuvent pas dévier une connexion vers un serveur sous leur contrôle.

Le logiciel client pour SSH existe sur toutes les plates-formes courantes, mais le logiciel du côté serveur n'existe à l'heure actuelle que pour Unix et Windows.
Les clients SSH offrent la possibilité de créer des tunnels SSH. A travers un menu ou par des options sur la ligne de commande (selon le client utilisé), on peut spécifier le serveur et le port (service) de ce serveur qui seront la cible du tunnel. Notons que ce serveur est en général différent du serveur SSH sur lequel le client a ouvert une session interactive. Le logiciel client SSH va alors ouvrir ce même port sur la machine cliente (il est possible de faire ouvrir un port différent, mais ceci est rarement utilisé dans la pratique). Quand une application se connecte à ce port, ouvert sur le client, le logiciel SSH sur le client et celui sur le serveur SSH collaborent pour que les données transmises sur ce port parviennent au serveur cible du tunnel, et inversement, les données émises par ce serveur parviennent à l'application connectée sur le port source du tunnel. L'utilisation des tunnels SSH passe par une connexion locale (sur le client) pour atteindre le serveur cible, exactement comme pour la connexion avec un serveur proxy « standard ».
Remarques diverses :

- Contrôles d'accès : en général, l'accès aux ports ouverts pour les tunnels est réservée au client SSH qui les a ouverts, pour des raisons de sécurité.
- Mentionnons également que le protocole SSH prévoit les tunnels inverses : un port est ouvert sur le serveur SSH (au lieu du client) et les données transmises par les applications à travers ce port arrivent à une application de type serveur sur le client SSH. Ceci n'est implémenté que sur le client SSH UNIX.
- Il y a toujours une session interactive ouverte sur le serveur SSH, même si la connexion SSH ne sert qu'à ouvrir des tunnels SSH.

De plus, ces données peuvent concerner aussi bien des e-mails, des articles de forums de discussion, des pages Web ou du trafic FTP.

Comment faire, pratiquement ? Ceci demande un logiciel spécifique, et surtout un serveur proxy qui accepte de chiffrer/déchiffrer les données. Ces services sont généralement payants, il y a par exemple le tunnel de Anonymizer ($30 pour 3 mois, http://www.anonymizer.com/services/ssh2.shtml).
Il est nécessaire de se procurer un logiciel[96] qui chiffrera/déchiffrera les données, ce dernier doit être compatible avec celui utilisé par le proxy. Il est nécessaire de configurer son logiciel[97] (en lui indiquant le nom DNS ou l'adresse IP du proxy) et lui signaler le type des données devant être traitées : e-mails, HTTP, FTP ou newsgroups. La paire de clés pour le chiffrage et le déchiffrage sont envoyées lors de la première connexion avec le proxy. Ces clés sont bien entendu personnelles et sont changées automatiquement à intervalles réguliers pour garantir une meilleure confidentialité.
Enfin, le logiciel (un browser Web, un logiciel de messagerie ou encore un programme de groupe de discussion) que l'internaute désire utiliser devra être configuré afin qu'il reconnaisse le tunnel et y envoie ses requêtes.

L'emploi d'un tunnel crypté demande donc un logiciel spécifique, et un paramétrage adéquat des logiciels employés. En contrepartie, toutes les données transitant par l'ordinateur de l'utilisateur seront cryptées et donc confidentielles. Et bien entendu, l'utilisateur bénéficie du masquage de son adresse IP qui est « remplacée » aux yeux des sites visités par celle du proxy.

b) Les Proxies Socks

Les proxies les plus courants, s'ils permettent de surfer en masquant l'adresse IP, ne seront pas efficaces, par exemple pour l'utilisation d'ICQ, les accès FTP ou encore certains services de Webmail. Par contre, les proxies de type « Socks » permettent tout cela. Avec un tel proxy, on peut « masquer » son adresse IP pour la plupart des activités en ligne possibles.
Disons juste que *Socks* (de type 4 ou 5) est un protocole[98] transparent de « proxying/tunneling » supporté par certains serveurs proxies (les « proxysocks » ou tout

---

[96] Comme, par exemple, les shareware SecureCRT (http://www.vandyke.com/products/securecrt/) ou F-Secure SSH (http://www.europe.f-secure.com), et les logiciels gratuits OpenSSH (http://www.openssh.com/) , TTSH qui est une extension spécifique pour Teraterm (http://www.zip.com.au/~roca/ttssh.html), ou Putty (http://www.chiark.greenend.org.uk/~sgtatham/putty/).

[97] Une fois le « client » SSH installé sur sa machine, la première chose à faire est d'aller dans les préférences et de désactiver les algorithmes de cryptage DES et 3DES (faibles et facilement crackables). On leur préférera les algorithmes IDEA, Blowfish, et RC4 (ou ArcFour), qui sont bien plus forts. Ce sont les seuls à véritablement sécuriser les connexions. Attention toutefois à la version du protocole utilisée... (Pour plus de renseignements, http://www.bugbrother.com/security.tao.ca/ssh-fr.htm ou http://www.employees.org/~satch/ssh/faq/ssh-faq.html)

[98] Pour de plus amples informations, http://www.socks.nec.com/aboutsocks.html. Il existe deux protocoles SOCKS : la version 4 (supportant uniquement TCP) et la version 5 (supportant TCP et UDP).
Les socks sont un moyen de faire passer les communications informatiques à travers une machine comme un proxy. On appelle ces dispositifs des relais. Parmi les relais, on distingue 2 types :

- les relais applicatifs qui comprennent tout le protocole qu'ils relaient, ce qui permet un suivi plus précis des transactions.
- les relais circuits qui se contentent de savoir quelle est la machine et le port d'origine, la machine et le port de destination. Ils font un contrôle à priori. La suite de la communication ne les concerne plus. Les socks sont des relais circuits.

simplement « Socks » par abus de langage), il permet de relayer la majorité des autres protocoles. Comme les proxies standard, ils permettent également à leurs utilisateurs de se connecter à Internet tout en étant anonyme vis-à-vis l'extérieur, seule l'adresse du proxy apparaissant sur le réseau.

Comme de nombreux proxies, les socks appartiennent en général à des réseaux d'organisations et sont normalement fermés au public... Dans la réalité, il en est autrement, et les opérateurs de certains de ces réseaux laissent l'accès libre... parfois jusqu'à ce qu'ils se rendent compte que leur proxy est utilisé par de tierces personnes.
On ne trouve donc que rarement des listes de Socks sur Internet, et celles qui existent doivent être mise à jour continuellement (voir tout de même http://proxys4all.cgi.net/win-tel-socks.shtml ou http://tools.rosinstrument.com/proxy/1080.htm).

Une fois que l'on a trouvé - et testé - un Socks, il suffit de dire au navigateur (ou à tout autre logiciel avec lequel on désire utiliser un service disponible sur Internet) de s'y connecter. Toutes les requêtes émises transiteront alors par ce Socks, ainsi que les réponses des serveurs destinataires.
En pratique, il est relativement fastidieux de paramétrer un par un chacun des logiciels que l'on veut utiliser. D'autant plus que certains s'y prêtent mal, voire pas du tout. Mais il existe un logiciel qui permet d'outrepasser ce problème tout en faisant transiter le trafic de pratiquement n'importe quelle application par un proxysocks : le logiciel gratuit (pour un usage non commercial) ***SocksCap*** (http://www.socks.nec.com/reference/sockscap.html). Ce logiciel intercepte et redirige vers le Socks choisi, tous les paquets (UDP et TCP) émis à partir de notre ordinateur, par les logiciels choisis. Il suffit donc de paramétrer uniquement SocksCap, sans toucher aux autres applications. Le jour où le Socks que l'on utilisait jusqu'à présent nous refuse une connexion, il suffit d'en indiquer un autre dans SocksCap, tout simplement. Ensuite, on exécute le logiciel de notre choix, non pas directement, mais à travers la fenêtre de SocksCap. Le paramétrage s'effectue en quelques minutes à peine, et l'utilisation de SocksCap est complètement transparente. Ce système permet de masquer son adresse IP lors de l'emploi d'ICQ, de logiciels FTP, de Telnet, et même de certains logiciels de News. SocksCap n'a qu'un défaut, que l'on retrouve dans toutes les applications faisant appel à des proxies : il ralentit notablement les connexions et les téléchargements. On peut également chaîner un Socks et un proxy traditionnel. Ainsi, si on envoit des requêtes par le Socks, qui les enverra ensuite au Proxy, qui les enverra enfin au serveur voulu, ce dernier ne verra bien entendu que l'adresse IP du proxy. Ceci permet de brouiller encore plus les pistes en cas de recherche…

Quelques précisions tout de même : il faut vérifier que le Socks utilisé ne transmette pas les adresses IP réelles de ses utilisateurs aux serveurs Web. Si c'est le cas, il vaut mieux changer de Socks. De plus, en théorie il est toujours possible de remonter à l'adresse IP de n'importe quel utilisateur, puisque la grande majorité des Socks disponibles appartiennent à des entreprises ou des administrations, celles-ci pouvant garder des fichiers log des connexions. En pratique, si on chaîne plusieurs Socks, il y a nettement moins de risques...

Comme on l'a vu, SocksCap redirige les connexions des applications pour lequel il est configuré vers un proxysocks par le biais du port 1080 - sans les modifier d'aucune manière - et garantit un anonymat quasi-parfait. Mais certains applications ne fonctionnent pas avec SocksCap : ce sont celles nécessitant l'envoi de multiples requêtes

---

Les proxysocks (type 4 ou 5) utilisent généralement le port 1080. Le Socks version 5 est décrit par la RFC 1928 (http://www.ietf.org/rfc/rfc1928.txt, cfr. [20]).

simultanées (SocksCap est limité à 50). De plus, pour une pleine compatibilité avec la majorité des applications, il faut utiliser de préférence un proxysocks de classe 5.

Le chaînage de socks permet donc de brouiller les pistes efficacement. Toute personne voulant « remonter la piste » devra contacter les administrateurs des Socks utilisés, consulter leurs logs, pour voir que la connexion vient d'un autre Socks. Pour peu que l'on chaîne des Socks originaire de certains pays (Malaysie, Russie, Afrique du Sud ou Chine), autant dire que l'on est quasi sûr de ne jamais être découvert.
En outre, il existe un logiciel permettant d'effectuer le chaînage de Socks de manière simple et automatique : ce logiciel s'appelle ***SocksChain*** (http://ufasoft.com/socks/, shareware, prix de $29, 2 semaines d'essai gratuit).

c) HTTPort

Le mode « Connect/SSL » est un mode de connexion[99], tout comme le mode « Get ». Mais le mode Connect, lorsqu'il est supporté par les proxies, permet de les utiliser également pour des applications autres que le HTTP. On peut alors poster des messages sur les forums, envoyer ou recevoir des e-mails, utiliser Telnet, faire de l'IRC, tout cela avec un proxy « classique », un peu comme si c'était un Socks (qui est aussi indépendant du protocole utilisé par les informations échangées). Quel avantage ? Les Socks sont relativement difficiles à trouver tandis que de nombreux proxies supportent le mode Connect.

Le logiciel HTTPort (logiciel gratuit, http://www.htthost.com/index.htm) permet d'utiliser ces proxies, pour la plupart des services existants (POP, SMTP, Telnet, IRC, HTTP, HTTPS, etc.). Plus simple d'utilisation que SocksCap et permettant, depuis sa version 3, l'utilisation des Socks, la fonction de base de HTTPort est l'interception de toutes les connexions effectuées par les logiciels qui permettent l'emploi des services désirés et la création d'une connexion à travers un proxy grâce à un tunnel SSL.

Avantage : HTTPort permet de passer au travers des proxies « obligés ». En effet, quand on travaille dans une entreprise, il est fréquent que toutes les connexions à Internet passent par l'intermédiaire d'un proxy paramétré de manière à enregistrer les sites qui sont visités par les employés et à interdire l'accès à certaines parties du réseau. Or HTTPort, grâce à son système de « tunneling » SSL, est capable de passer au travers du proxy institutionnel qui gère les connexions tout en préservant la possibilité de surfer anonymement.

II.5. Quelques mots sur les Firewalls personnels

Leur rôle : bloquer les accès indésirables venant d'Internet. L'installation d'un pare-feu personnel (càd un logiciel jouant le rôle de firewall et installé sur son propre ordinateur) n'est pas un luxe lorsqu'on dispose d'une connexion Internet haut débit et permanente (par exemple, lorsque l'on est relié à un modem câble ou à l'ADSL), c'est même un type de protection indispensable et très utile pour éviter de nombreux désagréments. Ces logiciels sont consacrés à la défense de l'intégrité de la machine sur laquelle ils sont installés en empêchant les attaques extérieures et en protégeant les échanges de données entre cette machine et le réseau Internet. Pour ce faire, ils épluchent les communications entrantes et sortantes de cette

[99] Voir http://developer.netscape.com/docs/manuals/proxy/ProxyUnx/SSL-TUNL.HTM pour un exposé détaillé de ce sujet.

machine en analysant les paquets IP qui circulent et en les filtrant. Lorsqu'ils détectent un accès entrant ou sortant suspect, ils peuvent laisser décider l'utilisateur sur le fait d'accepter ou non cet accès, pour une fois ou de manière systématique grâce à un ensemble de règles et de filtres. C'est aussi un bon moyen pour découvrir les programmes qui communiquent des données à l'insu de son utilisateur (voir ci-dessous). Certains pare-feu, comme Norton Internet Security, sont également capables de filtrer ou bloquer les bandeaux publicitaires.

A partir du moment où un ordinateur est connecté à Internet ou à un réseau local, il peut devenir l'objet de divers types d'attaques, abus ou autres actions indésirables. Les intrusions sont des actions qui sont accomplies pour compromettre l'intégrité, la confidentialité ou la disponibilité d'une ou plusieurs ressources et sont dépendantes de la politique de sécurité mise en œuvre sur cet ordinateur. Par conséquent, il est primordial que les internautes soient conscients qu'un ordinateur connecté à Internet est réellement vulnérable aux attaques potentielles de hackers : visites, vols, destructions de fichiers, introductions de virus (en particulier des chevaux de Troie) ou d'un code dommageable pour les données stockées. En effet, un ordinateur possède de nombreuses portes ouvertes, appelés ports, que les hackers vont tenter d'exploiter pour pénétrer dans le système informatique. Sachant qu'une machine dispose de 65536 ports pour communiquer avec Internet et que les ports inutilisés restent ouverts, ils représentent autant de portes d'entrées ! Chaque service Internet (ou protocole) à son numéro de port par défaut (par exemple, HTTP utilise le port 80, FTP le port 21, Telnet le port 23, le courrier expédié (SMTP) utilise le port 25 et celui reçu (POP3) le port 110).

De plus, Microsoft a choisi d'ouvrir des ports permettant l'échange de fichiers au sein d'un réseau dans Windows 9x/NT. Ces ports (par défaut les 137, 138, 139) sont de véritables trous béants pour les hackers malintentionnés qui ne cessent de les exploiter pour explorer les ordinateurs connectés au réseau comme autant de livres ouverts. Steve Gibson, informaticien spécialiste en sécurité, décrit une méthode, sur son site Web, permettant de s'assurer que le port 139 (et accessoirement les 137 et 138) ne restent pas ouverts et donc à la merci d'une éventuelle intrusion.

Le firewall ferme les ports non utilisés et peut parfois tenir à l'œil les applets Java et autres contrôles ActiveX. Quand un pare-feu détecte une connexion suspecte, il affiche une fenêtre d'alerte contenant plusieurs informations intéressantes dont l'adresse IP de l'intrus, certains identifient même son fournisseur d'accès Internet[100] ou permettent de riposter. La plupart sont capables de détecter un « scanning » des ports de l'ordinateur qui l'héberge, méthode assez courante pour détecter les failles dans un système (pour mieux l'attaquer ensuite).

En dehors des intrusions, le firewall assurera également la surveillance des informations émises par les « espiogiciels » (ou spyware), ces petits utilitaires (ou grandes applications) qui se connectent à Internet afin d'envoyer des informations personnelles à l'insu de l'utilisateur. Après avoir bloqué la connexion d'une application jugée douteuse, ZoneAlarm, par exemple, permet d'accepter ou de refuser (définitivement ou ponctuellement) tout accès à Internet ultérieur à cette application mais aussi pour n'importe quelle autre application ayant tenter d'accéder au réseau, même si elle ne représente aucun risque a priori.

[100] Notons au passage qu'il est toujours possible d'identifier la machine d'un intrus à partir de son adresse IP, grâce à la base d'interrogation WhoIs (http://www.ripe.net/cgi-bin/whois), sauf bien sûr si le pirate se cache derrière un proxy.

Le site de Steve Gibson, par exemple, permet de tester en ligne la sécurité de sa connexion Internet ainsi que tous les ports ouverts (test nommé « Shields Up !! » : https://grc.com/x/ne.dll?bh0bkyd2). Une bonne manière de mettre l'efficacité de son firewall à l'épreuve, un compte rendu très explicite étant généré avec des recommandations détaillées pour améliorer la situation (si toutefois un problème a été détecté). Voici quelques tests et diagnostics de sécurité supplémentaires :

http://check.sdv.fr/
http://www.pcflank.com/[101]
http://www.securityspace.com/sspace/index.html (Section Free Trial Audit)
http://www.dslreports.com/scan
http://www.blackcode.com/scan/
http://scan.sygate.com/

Local Port Scan est un petit logiciel gratuit conçu par la société danoise JPSoft (http://www.jpsoft.dk/) et qui permet de scanner les ports de son ordinateur afin de vérifier s'ils sont ouverts ou fermés. Plusieurs modes de scan sont proposés : « Trojan Scan » (vérification des ports utilisés par les Chevaux de Troie les plus communs), « Stealth Scan » (vérification de la « discrétion » de son ordinateur en ligne), « Full TCP Scan » (scan des ports 1 à 1024), et quelques autres encore. LPS n'est pas parfait (mais rien ne saurait remplacer un scan systématique pratiqué avec un ordinateur interconnecté directement), car il ne passe pas en revue tous les ports susceptibles d'être « détournés ». Néanmoins, il est très simple d'usage et instructif (il fournit, par exemple, des informations concernant les ports vulnérables aux Trojans les plus répandus), et offre également la possibilité de créer soi-même de nouveaux types de scans. Pour que le test soit positif, il faut voir la mention « Blocked » apparaître à côté de chaque numéro de port testé.

Heureusement, de nombreux firewalls s'adressent aujourd'hui au grand public et sont relativement simples à configurer. Le produit suivant en est un exemple puisqu'il utilise une solution logicielle facile à mettre en oeuvre pour appliquer des mesures de sécurité efficaces à un PC sous MS Windows. ***Norton Internet Security 2002*** (shareware, $69.95) est un logiciel publié par Norton (http://www.norton.com). Il donne à son utilisateur l'opportunité d'établir un véritable firewall personnel sur son propre ordinateur, qui protégera ce dernier contre les intrusions de personnes tentant de l'utiliser pour leurs propres besoins ou d'obtenir des informations confidentielles. Le programme accomplit également plusieurs autres tâches plus spécifiques à la protection de la vie privée, comme le blocage des bannières publicitaires et la gestion des cookies. Quand un utilisateur emploie ce logiciel pour installer son propre firewall, il rend son ordinateur « invisible » sur Internet et en principe, inaccessible aux hackers. Les firewalls personnels adoptent les mêmes technologies que les grandes entreprises emploient pour protéger leurs réseaux tout en les rendant plus abordables et faciles à utiliser par les novices.
L'utilisateur peut demander au logiciel de refuser la connexion de certaines applications à Internet[102], mais il peut aussi éditer ses propres règles (contrôle des ports et des protocoles

[101] *PCflank* est devenu un site de référence en ce qui concerne les tests de sécurité en ligne. Complet, très clair (bien qu'en anglais), il offre un grand choix de tests utiles : discrétion ("stealth)", résistance aux « exploits » (faiblesses inhérentes aux systèmes), ports (configurable), etc. A essayer absolument !

[102] Ainsi, l'utilisateur peut décider de bloquer la capacité de faire des achats en ligne à des personnes utilisant son ordinateur si elles n'ont pas sa permission. Il peut aussi donner des instructions au logiciel pour interdire certains types d'activités sur Internet pour divers utilisateurs, par exemple si une famille partage le même ordinateur, les parents pourraient ainsi empêcher leurs enfants d'aller consulter les « chat rooms ».

utilisés) afin d'avoir un contrôle précis des informations échangées avec Internet. En outre, cette solution logicielle inclut désormais un programme antivirus (celui de l'éditeur) et un système de contrôle parental.
L'utilisateur a la possibilité de donner des instructions au logiciel pour qu'il refuse la diffusion de certains fichiers ou certaines informations que l'on désigne comme confidentiels. Lorsqu'une donnée personnelle déterminée est désignée, le programme interdira quiconque utilisant l'ordinateur sur lequel il s'exécute de distribuer ces informations à de tierces personnes au moyen d'Internet. L'utilisateur peut aussi spécifier différents réglages de sécurité pour différents sites Web selon la confiance qu'il leur accorde.
Signalons également que d'autres éditeurs publient des produits similaires tout aussi efficaces au niveau des protections mises en place mais un peu moins complet au niveau de certaines fonctionnalités et qui ont le grand avantage d'être totalement gratuits. Par exemple, les firewalls populaires ***ZoneAlarm*** (http://www.zonelabs.com/) ou ***Sygate Personal Firewall*** (http://www.sygate.com/) sont gratuits pour une utilisation dans le cadre privé et tout à fait accessibles aux novices[103]. ***Tiny Personal Firewall*** (shareware, la version 2 est gratuite pour un usage privé, http://www.tinysoftware.com/), ***VisNetic Firewall*** (shareware, http://www.ccsoftware.ca/, remplace l'ancien Conseal PC Firewall) et ***Looknstop*** (shareware existant en version « Lite » gratuite, http://www.looknstop.com/) sont trois firewalls très puissants réservés aux utilisateurs expérimentés désirant une sécurité optimale pour leur ordinateur (à condition qu'ils sachent les configurer correctement).

II.6. Conclusion

Le but de ce deuxième chapitre était de faire un large tour d'horizon des acteurs en présence. La « destruction » de la vie privée est aujourd'hui la préoccupation majeure de nombreux internautes. Le débat dépasse désormais le cadre de l'Internet car la problématique de la « privacy » occupe désormais une place importante dans le débat américain ou européen. Pour une grande part de l'opinion de la communauté des internautes, la destruction de la vie privée devient à l'économie de l'information ce que la destruction de l'environnement était à l'économie industrielle : une pollution. Rien d'étonnant dès lors à ce qu'apparaisse un nouveau marché d'entreprises de « dépollution », dont le modèle économique repose souvent sur la participation financière des internautes. Aujourd'hui, de nombreuses sociétés peuvent être citées comme acteurs de ce marché émergent des services d'anonymisation et de démocratisation de la cryptographie. Elles ont en général la particularité d'être spécialisées sur un outil technologique particulier. Ainsi, de nouveaux acteurs font aujourd'hui le pari que les services de e-Privacy ne sont pas les adversaires des approches de cybermarketing mais bien une de leurs composantes.

Bien sûr, rien n'empêche un internaute d'utiliser simultanément plusieurs technologies de protection des données à caractère personnel. Ainsi, il pourrait utiliser conjointement :

- les outils P3P pour connaître les politiques des sites qu'il visite en matière de vie privée et de les comparer avec ses propres préférences,
- les labels Internet pour s'assurer que les sites satisfont bien leurs politiques,
- les logiciels ou services d'anonymisation et/ou de filtrage pour réduire la quantité d'information révélée lors de sa navigation,

[103] On peut trouver de nombreuses informations, en français, sur les firewalls les plus répandus auprès du grand public sur le site http://www.firewall-net.com/.

- les logiciels de chiffrage pour sécuriser les échanges de données confidentielles ou sensibles,
- la législation et les documentations, disponibles sur Internet ou ailleurs, qui concernent le respect de la vie privée pour se faire une idée du contenu que doit posséder une politique pour être « acceptable » et avoir un regard critique vis-à-vis des déclarations en matière de protection de la vie privée des sites qu'il fréquente.

Malheureusement, les motivations précises de toutes les technologies de renforcement de la protection de la vie privée que nous avons évoquées tout au long de ce chapitre ne sont pas forcément « évidentes » pour un novice découvrant le monde d'Internet.
Il est important pour les utilisateurs de comprendre l'ampleur des processus et techniques de collecte de données à caractère personnel ainsi que leur capacités de profilage. Mais il doit également comprendre les capacités qu'il peut avoir pour limiter la quantité d'information qui est récoltée pendant qu'il profite des services offerts par Internet. Bien sûr, la collecte commence au niveau du fournisseur d'accès à Internet, puisque celui-ci est capable de répertorier tous les sites Web demandés par chacun de ses clients pendant leurs sessions en ligne mais aussi l'origine, la destination et le contenu des e-mails envoyés par ses utilisateurs vers le monde extérieur. La conscience de ces capacités devrait rappeler constamment aux internautes que leurs actions sur Internet sont loin d'être anonymes et peuvent même être utilisées contre eux bien après qu'ils aient supposés que les informations récoltées à leur sujet ont été supprimées définitivement.
Certains internautes informés veulent donc maîtriser cette collecte pour préserver leur vie privée. D'autres veulent simplement bloquer les bannières publicitaires qui les embarrassent et encombrent les pages Web ou les fenêtres pop-up qui apparaissent intempestivement lors de leur navigation.

Différentes organisations se sont donc emparés de ce marché et ont commencé à porter leur attention sur le développement de solutions à ces questions. Ces organisations ont produit des logiciels pour neutraliser les menaces à la vie privée. Cependant, de nombreux consommateurs ignorent les types d'outils de protection de la vie privée disponibles sur le marché. C'est pourquoi ce chapitre essaie de faire un bilan de l'éventail des produits accessibles aux consommateurs individuels. Finalement, nous pouvons classer ces produits en trois catégories primaires :

1. **Mécanismes d'Anonymat** qui protégent et masquent l'identité de l'utilisateur,
2. **Mécanismes de Filtrage** qui empêchent (ou permettent de contrôler) l'affichage de publicités non sollicitées, la diffusion d'informations (comme les cookies ou le bavardage des navigateurs), l'utilisation de contenus actifs potentiellement dangereux (scripting côté client malicieux), la réception d'e-mails non souhaités (« Spam ») et d'autres objets indésirables,
3. **Mécanismes de Choix** qui améliorent la maîtrise de l'internaute sur le contenu qu'il veut ou non consulter.

Il serait assurément préférable que les éditeurs de logiciels conçoivent des navigateurs conformes au respect de la vie privée, et dont la configuration par défaut est plus favorable à la protection de la vie privée.

Voici quelques points qui amélioreraient significativement le respect de la vie privée des internautes s'ils étaient suivis par les plus importants acteurs du marché (éditeurs et fournisseurs d'accès à Internet, notamment) :

- concernant les serveurs proxies anonymes permettant de dissimuler l'adresse IP, ils devraient être proposés gratuitement en standard lors de chaque inscription Internet auprès d'un fournisseur d'accès (ISP) ;
- les sites Web ne devraient pas empêcher leur accès à des utilisateurs qui refusent les cookies, à moins que des cookies de session ne soient indispensables pour, par exemple, faire un lien entre l'utilisateur et son « identité » en ligne, et donc permettre de faire du e-commerce (cfr. [9]), par exemple ;
- il semble nécessaire de mieux informer les particuliers à propos de l'existence de technologies de renforcement de la protection de la vie privée (PET). Le secteur public, ainsi que les éditeurs des navigateurs, devraient prendre les mesures nécessaires pour augmenter la notoriété de ces instruments, en soutenir le développement, les utiliser et en promouvoir l'emploi ;
- il faudrait encourager l'utilisation de PET, surtout s'ils sont installés par des ISP ou d'autres acteurs ;
- des normes européennes pour les labels de protection de la vie privée pourraient être conçues. Ces normes devraient inclure l'obligation faite aux sites Web de se soumettre à des audits périodiques.

Chapitre III : Spécification du ProxyServeur et Présentation du Filtre réalisé pour Jigsaw

Nous avons constaté dans le deuxième chapitre que les browsers les plus répandus laissent passer un certain nombre d'informations à caractère personnel concernant leurs utilisateurs sans que ces derniers en soient informés. Il convient donc, puisque le but du mémoire consiste aussi à réaliser un ProxyServeur d'Anonymisation, de choisir quelle est la manière la plus appropriée, donc la plus respectueuse de la vie privée, pour le proxy que nous voulons concevoir de contrôler ces « fuites » d'informations en rapport avec les différents risques potentiels que nous avons mis en lumière au premier chapitre. Il est d'ores et déjà utile de préciser que ce sera l'ensemble des mesures prises à l'encontre de ces risques qui constituera une réelle solution de protection. En effet, dans le premier chapitre, nous nous sommes rendus compte qu'il ne fallait pas considérer uniquement chaque risque pris séparément, puisque c'est en les utilisant conjointement que les sociétés de marketing parviennent à établir le plus efficacement possible les profils des internautes.

Ce dernier chapitre exposera les différentes mesures protectrices, ainsi que les différentes configurations possibles, que peut prendre le proxyserveur que nous avons finalement conçu. Nous motiverons finalement les choix que nous avons fait quant à la réalisation proprement dite du proxyserveur d'anonymisation et décrirons brièvement notre travail d'implémentation.

III.1. Mesures protectrices et configurations du ProxyServeur d'Anonymisation

III.1.1. Communication de l'adresse IP

Concernant la communication de l'adresse IP par les protocoles TCP/IP, le proxyserveur (non transparent) que nous désirons mettre au point aura comme fonction de cacher l'adresse de ses utilisateurs en utilisant sa propre adresse IP pour communiquer avec Internet, et sans divulguer celles de ses utilisateurs (ni aucune autre information qui dévoilerait leurs présences derrière le proxy), ainsi ceux-ci ne communiquent pas directement leurs « identités » informatiques sur le réseau mondial. De plus, le proxy doit se comporter comme un client à part entière du point de vue du serveur Web avec lequel un utilisateur souhaite rentrer en contact, et donc ne pas déclarer sa véritable nature (càd un proxy servant les requêtes de plusieurs clients).

III.1.2. Bavardage des browsers

Comme les browsers sont « très bavards », il serait adéquat de limiter ce bavardage. Par exemple, il serait utile de modifier le champ déclarant le navigateur et le système d'exploitation (et leur version) utilisés par l'internaute. Ainsi, nous pensons qu'il serait utile d'indiquer en lieu et place la signature d'un autre navigateur (par défaut, un navigateur assez ancien afin que ses fonctionnalités soient supportées par tous les navigateurs des utilisateurs

potentiels[104] ; bien sûr, le choix que nous avons effectué peut être remplacé selon la volonté de l'administrateur du proxy). Cette fausse identification aide à résister à certaines attaques.
Il est certainement plus pertinent de supprimer le champ HTTP_ACCEPT_LANGUAGE déclarant les langues acceptées par le browser (car il est totalement inutile pour la navigation).
Il en est de même pour le champ HTTP_REFERER indiquant la page référente car cette information joue un rôle important (bien qu'elle puisse être subtilement contournée) dans le processus de profilage d'un internaute par une société de cybermarketing, ainsi que pour les champs HTTP_FROM et HTTP_VIA (le premier pouvant indiquer, s'il est communiqué par le browser, l'adresse e-mail de son utilisateur tandis que le deuxième dévoile la présence d'un proxy devant celui-ci).
Enfin, nous pourrions mettre le champ HTTP_ACCEPT indiquant les types de fichiers supportés par un navigateur à une valeur par défaut « passe partout », en l'occurrence la valeur */*.
Ces cinq dernières options doivent bien sûr être désactivables et la signature du browser communiquée par le proxy doit être paramétrable.

III.1.3. Hyperliens invisibles et Redirection automatique

Comme nous l'avons signalé au cours du premier chapitre, le protocole HTTP est « stateless » (voir section I.3.2). Ceci implique qu'un proxy recevant et transmettant les requêtes d'un de ses clients ignorent totalement si cette requête émane d'un hyperlien visible ou invisible. Ne pouvant faire la distinction, il ne peut pas bloquer sélectivement l'un ou l'autre. Il faut donc trouver une autre solution pour résoudre le risque correspondant.

D'autre part, pour se prémunir du risque de la redirection automatique, nous pourrions bloquer systématiquement toutes les réponses indiquant une redirection et répondre aux clients par une réponse HTTP indiquant une erreur. Cette solution n'est tout fois pas envisageable pratiquement puisque de nombreux sites utilisent désormais la redirection pour rediriger leurs visiteurs vers une autre page lorsque la page demandée ne se trouve pas à l'URL indiquée. Par conséquent, il faut trouver une autre solution moins grossière.

Pour résoudre les deux problèmes évoqués dans les deux précédents paragraphes, nous avons préféré opter pour une solution du type « liste noire d'URLs » telle qu'elle est utilisée aujourd'hui par plusieurs logiciels de protection des données à caractère personnel (Junkbuster notamment). Nous avons fait en sorte qu'elle soit la plus flexible possible afin de rendre la tâche de son « remplissage » la moins fastidieuse possible. Il serait en effet pénible de devoir indiquer un à un chaque site que l'administrateur désirerait bloquer au nom de la protection de ses clients. Heureusement, il existe, sur Internet, des fichiers *Hosts* (voir II.4.3.b) judicieusement complétés qu'il suffit de modifier légèrement pour créer une telle liste noire. Bien sûr, cette liste peut être mise à jour à n'importe quel moment, sans devoir redémarrer le logiciel. De plus, son utilisation est désactivable.

III.1.4. Cookies

En ce qui concerne les cookies et leur gestion, la solution la plus pertinente (càd la configuration par défaut) aurait été la suivante : bloquer les cookies de sites tiers avec la possibilité de toujours retransmettre les cookies de session. Seulement, comment faire la

[104] Rappelons que le proxyserveur pourra être utilisé par un nombre quelconque d'utilisateurs et que la nouvelle signature de navigateur remplacera toutes celles des browsers employés par ces derniers.

distinction entre les cookies tiers et les cookies dont l'origine est le site visité ? En effet, chaque requête ne contient jamais que des cookies dont le domaine est un « sous-ensemble » du domaine de l'URL indiquée et comme une requête ne peut être distinguée des autres, comment reconnaître un site tiers d'un autre (il s'agit ici du même problème que celui précédemment rencontré avec les hyperliens invisibles et dû au fait que le protocole HTTP est sans état). N'oublions pas qu'un proxy ne se base que sur le contenu des en-têtes des requêtes HTTP qui lui sont transmises et n'interprètent pas le contenu de ces requêtes (il n'est pas un navigateur, par conséquent il ne « comprend » pas le HTML). Ainsi, seul un navigateur sur le poste client est capable de faire la distinction entre un site tiers et un site « normal ».

Nous avons donc dû procéder à d'autres choix. En configuration par défaut, le proxyserveur devrait gérer lui-même les cookies pour le compte de tous ses clients (ceux-ci n'en reçoivent donc plus). De plus, il n'acceptera que les cookies de session, tous les autres cookies seront supprimés dès leur arrivée au proxy. Il devra également être capable de gérer une double liste contenant l'origine des cookies qu'il doit bloquer sans distinction (même si ce sont des cookies de session), la « liste noire », et ceux qu'il doit accepter sans distinction ni modification, la « liste blanche ».

Par contre, on peut judicieusement offrir différentes options supplémentaires à l'utilisateur (administrateur) du proxyserveur (configurations personnalisées) :

a- bloquer tous les cookies ou tous les laisser passer (pourquoi pas ?) dans un sens ou dans un autre (sens Client -> Serveur ou sens Serveur -> Client), ou dans les deux sens, au choix.

b- bloquer ou laisser passer les cookies de session quelles que soit les options concernant les cookies persistants.

c- activer/désactiver la gestion des cookies provenant de sites Web distincts en désignant explicitement ceux dont on désire bloquer ou laisser passer systématiquement les cookies (liste noire et liste blanche, qu'il faut bien entendu compléter manuellement).

d- modifier les cookies persistants en permettant à l'administrateur du proxy de fixer lui-même la durée de vie maximale de tous les cookies qui traversent le proxy (durée que l'on définirait en jours) et activer/désactiver cette fonctionnalité.

e- activer/désactiver la fonction de gestion des cookies du proxy pour ses clients.

Bien sûr, l'administrateur pourra également choisir l'emplacement des fichiers texte contenant les différentes listes (notamment celle qui contient les cookies reçus et gérés par le proxy au nom de ses clients).

III.1.5. Scripting côté client

Signalons tout d'abord que le but du proxyserveur que nous voulons réaliser est de permettre à ses (futurs) administrateurs de protéger les clients des exemplaires du proxyserveur qu'ils administrent (y compris eux-mêmes, pourquoi pas ?) contre les risques les plus généraux d'atteinte à la vie privée (càd les risques « standards », non spécifiques), et non contre les risques plus particuliers (càd spécifiques dans le sens où ils dépendraient de plusieurs facteurs différents et particuliers). Ainsi, le scripting côté client est un risque que nous pouvons qualifier de spécifique puisqu'une faille de sécurité est, la plupart du temps,

propre à telle version de tel OS, à tel navigateur, à telle version de Java installée,... et donc dépend de plusieurs paramètres bien distincts qui ne sont pas réunis simultanément par une grande majorité de configurations d'ordinateurs utilisés par les clients potentiels.

Signalons également que, de plus en plus, la protection (pas encore réellement efficace) contre le type de menaces potentielles représenté par le scripting côté client est prise en charge par un autre type de logiciels, les AntiVirus.

Par conséquent, la seule protection que le proxyserveur prend vis-à-vis de ce risque est la fausse identification donnée par le côté client du proxy (qui fait croire aux sites Web qu'il est un ancien navigateur). Ainsi, un hacker malintentionné utilisera les mauvais « exploits » pour attaquer un utilisateur caché derrière le proxy.

III.1.6. Mémoire cache

Le proxyserveur doit pouvoir offrir une certaine flexibilité en ce qui concerne le cache. En effet, si c'est un particulier qui utilise le proxy pour son usage personnel et exclusif, pourquoi n'activerait-il pas la mémoire cache de ce logiciel pour pouvoir bénéficier des avantages que cette fonctionnalité procure.
Par contre, si c'est une société ou un ISP qui administre le proxyserveur pour ses employés ou ses clients, il devrait pouvoir désactiver cette fonction pour garantir la vie privée de ses utilisateurs. Mais si cette société désire tout de même profiter des atouts de la mémoire cache tout en préservant au maximum la vie privée de ses employés, elle pourrait activer le partage de cette mémoire entre tous les utilisateurs afin d'être incapable de retracer le parcours d'un utilisateur particulier.

III.2. Jigsaw et le PrivacyProtectionFilter

Le but de ce mémoire est de créer un proxyserveur répondant au « cahier des charges » (ou spécification) exposés dans la section III.1 ci-dessus. Pour ce faire, nous avons opté pour la création d'un filtre côté client permettant d'implémenter les mesures de protection choisies au sein du serveur Web **Jigsaw** mis au point par le W3C (http://www.w3.org/jigsaw).
En effet, plusieurs possibilités s'offraient à nous :

1- développer un proxyserveur de A à Z
2- partir d'un proxy existant et le modifier de façon appropriée

Il va de soi que la première solution était irréalisable en pratique, dans le cadre de la réalisation d'un mémoire. En effet, une telle possibilité exigerait notamment l'implémentation complète du protocole HTTP/1.1 ce qui est déjà une tâche fastidieuse. La deuxième solution était donc la seule réellement envisageable dans le cadre de ce mémoire. Restait à savoir quel proxy préexistant allait-on choisir. Il nous fallait bien sûr trouver un logiciel disponible en open source et dont la licence permettait la réalisation de notre projet. Nous avions également une préférence pour un projet développé en Java puisque le langage Java est celui que nous maîtrisons le mieux. Toutes ces contraintes ne laissaient pas beaucoup de choix et nous avons opté directement pour le serveur Jigsaw mentionné supra, puisqu'il peut aisément être configuré en proxyserveur et bénéficie d'une API complète pour permettre de développer des extensions.

Pour une présentation intégrale de ce logiciel, nous invitons le lecteur à consulter la documentation complète le concernant (voir référence [25] et [26]).
Une fois le choix de la spécification et du logiciel effectué, il ne restait plus qu'à implémenter l'extension permettant d'étendre Jigsaw afin qu'il réponde à la spécification énoncée.

Le lecteur pourra trouver dans l'Annexe B de ce mémoire l'ensemble du code source, écrit en Java, réalisant les objectifs énoncés.
En l'occurrence, il s'agit du package *org.w3c.www.protocol.http.privacy* contenant les classes suivantes :

- *org.w3c.www.protocol.http.privacy.PrivacyProtectionFilter* (classe principale implémentant le filtre de protection des données à caractère personnel, elle « attrape au passage » les requêtes et les réponses HTTP reçues par le proxy et les gère/modifie selon sa configuration)

- *org.w3c.www.protocol.http.privacy.StringFromFile* (classe récupérant sous forme de String le contenu d'un fichier texte)

- *org.w3c.www.protocol.http.privacy.BlackList* (classe implémentant la liste noire d'URLs, correspondant aux ressources Web bloquées par le proxy)

- *org.w3c.www.protocol.http.privacy.DomainTree* (classe permettant au proxy de gérer les cookies pour le compte de ses clients)

- *org.w3c.www.protocol.http.privacy.CookieLists* (classe implémentant la liste noire (respectivement blanche) des domaines de cookie refusés (respectivement acceptés sans modification))

De plus, afin d'intégrer à Jigsaw le package réalisé et d'offrir une interface utilisateur permettant de configurer le filtre *PrivacyProtectionFilter*, il était nécessaire de modifier la classe suivante, située dans le package *org.w3c.jigsaw.proxy* :

- *org.w3c.jigsaw.proxy.ProxyProp* (classe permettant de paramétrer les options de Jigsaw lorsqu'il est configuré pour fonctionner en tant que proxy, la modification apportée permet d'ajouter le filtre *org.w3c.www.protocol.http.privacy.PrivacyProtectionFilter*)

et d'ajouter la classe suivante, située dans le même package que la classe précédente :

- *org.w3c.jigsaw.proxy.PrivacyProtectionProp* (classe permettant de modifier la configuration par défaut du filtre *org.w3c.www.protocol.http.privacy.PrivacyProtectionFilter* au moyen d'une interface graphique)

Pour de plus amples informations sur ces différentes classes, leurs packages, leurs méthodes et leurs paramètres, nous invitons le lecteur à consulter la JavaDoc[105].

[105] Il s'agit d'un ensemble de pages HTML décrivant en détail les classes présentées ci-dessus. Cette documentation, le code source et les classes compilées se situent sur la disquette remise avec le manuscrit.

Conclusion

La réalisation de ce mémoire nous a conduit à remplir un double objectif.
D'une part, nous avons évoqué l'existence de traces laissées par les internautes, à leur insu, pendant leur navigation sur le Web, ces informations pouvant alors être « récoltées » et exploitées par des tiers. D'où, l'étude des différents risques concernant la vie privée qu'encourent les internautes et la démonstration, par l'exemple, de l'efficacité des techniques de collecte des données à caractère personnel lorsqu'elles sont utilisées conjointement. Rappelons également que ces dangers potentiels puisent leur origine dans les fondements des nouvelles technologies de l'information et de la communication.
L'internaute n'a alors pas d'autres choix, s'il ne veut pas faire l'objet d'un profilage à grande échelle, que de s'informer et de trouver des parades. C'est là qu'intervient une nouvelle génération d'organisations et d'associations d'internautes afin de créer des « remèdes » adaptés aux différents besoins de protection des internautes. Ce qui nous a amené à présenter un large tour d'horizon et une présentation succincte des produits relatifs à la protection de la vie privée et qui sont disponibles sur le marché à l'heure actuelle. Nous avons pu distinguer essentiellement trois catégories distinctes parmi les logiciels/services présentés. Un internaute peut bien entendu cumuler l'usage de produits de catégories différentes pour renforcer davantage la protection de sa vie privée et la sécurité/confidentialité de ses échanges d'informations avec l'extérieur. Malheureusement, les internautes devront souvent mettre la main au portefeuille pour assurer leur propre protection tandis que les navigateurs qu'ils utilisent sont disponibles gratuitement. De plus, ils devront avoir une confiance aveugle envers les sociétés dont ils louent les services.
Cependant, il est nécessaire de mieux informer les internautes à propos de l'existence de technologies de renforcement de la protection de la vie privée dont il faut encourager l'utilisation, de préférence gratuite.

D'autre part, nous avons eu comme ligne directrice la réalisation, par nos soins, d'un produit libre, ouvert et efficient permettant à ses (futurs) utilisateurs de se prémunir des risques les plus génériques en matière de vie privée. Pour ce faire, nous avons choisi d'implémenter un proxyserveur disposant des filtres nécessaires à l'exécution de notre projet. Nous pensons avoir rempli notre tâche puisque le proxyserveur choisi (en l'occurrence Jigsaw créé par le W3C), lorsque ses fonctionnalités sont étendues par le filtre que nous avons conçu, satisfait les conditions minimales que devrait remplir tout proxyserveur d'anonymisation digne de ce nom. Bien sûr, notre réalisation peut encore être largement améliorée, les possibilités d'évolution étant nombreuses. Dans l'avenir, nous pourrions envisager des fonctions supplémentaires pour, par exemple, interpréter le langage HTML afin de pouvoir faire la distinction entre site visité et site tiers ou pour désactiver l'utilisation du JavaScript dans les pages Web.

BIBLIOGRAPHIE

[1] Jean-Marc Dinant, Chercheur au Centre de Recherches Informatique et Droit (CRID), Consultant auprès de la Commission Belge de protection des données et Expert auprès du Groupe de travail 29 sur la protection des données, « Les traitements invisibles sur Internet », école d'été à l'Université du Luxembourg, juillet 1998, Cahiers du CRID, n°16, 1999, p. 277-.302, Bruylant, Namur
http://www.droit.fundp.ac.be/crid/eclip/luxembourg.html

[2] Jean-Marc Dinant, Chercheur au Centre de Recherches Informatique et Droit (CRID), Consultant auprès de la Commission Belge de protection des données et Expert auprès du Groupe de travail 29 sur la protection des données, « Le visiteur visité - Quand les éditeurs de logiciel Internet passent subrepticement à travers les mailles du filet juridique », Lex Electronica, vol. 6, n°2, hiver 2001
http://www.lex-electronica.org/articles/v6-2/dinant.htm

[3] Jean-Marc Dinant, Chercheur au Centre de Recherches Informatique et Droit (CRID), Consultant auprès de la Commission Belge de protection des données et Expert auprès du Groupe de travail 29 sur la protection des données, « Faut-il renoncer à sa vie privée pour surfer sur le Web ? », Août 1999
http://www.droit.fundp.ac.be/Textes/revuegeneralejmd.rtf

[4] Jean-Marc Dinant, Chercheur au Centre de Recherches Informatique et Droit (CRID), Consultant auprès de la Commission Belge de protection des données et Expert auprès du Groupe de travail 29 sur la protection des données, « Vie privée, cybermarketing et cryptographie », Août 1999

[5] Jean-Marc Dinant, Chercheur au Centre de Recherches Informatique et Droit (CRID), Consultant auprès de la Commission Belge de protection des données et Expert auprès du Groupe de travail 29 sur la protection des données, « Law and Technology Convergence in the Data Protection Field - Electronic threats on personal data and electronic data protection on the Internet », ESPRIT Project 27028, Electronic Commerce Legal Issues Platform, Novembre 1999
http://www.jura.uni-muenster.de/eclip/documents/deliverable_2_2_3_privacy.pdf

[6] Groupe sur la protection des personnes à l'égard du traitement des données à caractère personnel (Working Party 37 du Groupe 9), CE, « Une approche européenne intégrée sur la protection des données en ligne », Document de travail, Le respect de la vie privée sur Internet, 21 novembre 2000
http://europa.eu.int/comm/internal_market/fr/media/dataprot/wpdocs/index.htm

[7] Groupe d'experts sur la sécurité de l'information et la vie privée, Direction de la Science, de la Technologie et de l'Industrie, Comité de la Politique de l'Information, de l'Informatique et des communications, Organisation de Coopération et de Développement Economiques, « Mise en oeuvre dans l'environnement électronique, et en particulier sur Internet, des lignes directrices de l'OCDE sur la protection de la vie privée », Septembre 1998
http://www.oecd.org/pdf/M000014000/M00014554.pdf

[8] Serge Gauthronet & Frédéric Nathan, ARETE - Coopérative Informatique, « Les services en ligne et la protection des données et de la vie privée - Etude pour la Commission des Communautés Européennes (DG XV) », Décembre 1998
http://europa.eu.int/comm/internal_market/en/dataprot/studies/servfr.pdf

[9] Thierry Léonard, Avocat au Barreau de Bruxelles, Assistant à la Faculté de Droit de Namur (Centre de Recherche Informatique et Droit), « E-Commerce et Protection des Données à caractère personnel - Quelques considérations sur la licéité des pratiques nouvelles de marketing sur Internet », Février 2000
http://www.droit.fundp.ac.be/Textes/Leonard1.pdf

[10] Pierre Barthelemy, Laboratoire de valorisation de l'information, Faculté des Sciences de Saint-Jérome, « Fiabilité de l'Information et Fragilité de l'Internet », Octobre 2001
http://iml.univ-mrs.fr/ftp/barthelemy/DESS/fiabilite.pdf

[11] Commission Nationale de l'Informatique et des Libertés
Service Information-Documentation, « Vos Traces sur Internet - Découvrez comment vous êtes pistés sur Internet », Octobre 1998
http://www.cnil.fr/traces/vostraces1a.pdf

[12] Conseil de l'Europe, Comité des Ministres, Recommandation N° R (99) 5 du Comité des Ministres aux Etats Membres sur la Protection de la Vie Privée sur Internet, « Lignes Directrices pour la protection des personnes à l'égard de la collecte et du traitement de données à caractère personnel sur les "inforoutes" », Février 1999
http://cm.coe.int/ta/rec/1999/f99r5.htm

[13] Dr. Lorrie Faith Cranor, AT&T Labs-Research, « Online Privacy Technologies », NTIA Online Privacy Technologies Workshop, Septembre 2000
http://lorrie.cranor.org/privacy/NTIA-0900.ppt

[14] Cyberspace Policy Institute, The George Washington University, School of Engineering and Applied Science, « Managing Your Identity on the Internet : Towards A Consumer Guide to Privacy Tools », Draft, Novembre 2000
http://www.privacyexchange.org/gpd/sites/privacytoolsreport.pdf

[15] Information Technology Industry Council, « Personal Privacy Solutions - A Look at Privacy Enhancing Technologies Available to Consumers », Juin 2000
http://www.itic.org/digital_frontier/consumer/privtech.pdf

[16] Arnaud BELLEIL, Directeur Marketing de Cecurity.com et Enseignant à l'IEP de Rennes, « Panorama des Entreprises Américaines de e-Privacy », Novembre 2000
http://www.fing.org/ref/identite/ePrivacyUS-Belleil.doc

[17] Ian Avrum Goldberg, « A Pseudonymous Communications Infrastructure for the Internet », A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Computer Science in the Graduate Division of the University of California at Berkeley, Automne 2000
http://www.isaac.cs.berkeley.edu/~iang/thesis-final.pdf

[18] Jonathan M. Smith, Chief R&D Consultant, iPrivacy.com, « A Brief Introduction to iPrivacy », Septembre 2001
http://www.paris-conference-2001.org/eng/contribution/smith_contrib.pdf

[19] Micheal K. Reiter, Aviel D. Rubin, AT&T Labs-Research, « Crowds : Anonymity for Web Transactions », Avril 1999
http://www.research.att.com/projects/crowds/papers/i8.pdf

[20] M. Leech, Bell-Northern Research Ltd, M. Ganis, International Business Machines, Y. Lee, NEC Systems Laboratory, R. Kuris, Unify Corporation, D. Koblas, Independent Consultant, L. Jones, Hewlett-Packard Company, « SOCKS Protocol Version 5 », Request for Comments 1928, Mars 1996
http://www.ietf.org/rfc/rfc1928

[21] R. Fielding, UC Irvine, J. Gettys, J. Mogul, DEC, H. Frystyk, T. Berners-Lee, MIT/LCS, « Hypertext Transfer Protocol -- HTTP/1.1 », Request for Comments 2068, Janvier 1997
http://www.ietf.org/rfc/rfc2068

[22] D. Kristol, Bell Laboratories, Lucent Technologies, L. Montulli, Netscape Communications, « HTTP State Management Mechanism », Request for Comments 2109, Février 1997
http://www.ietf.org/rfc/rfc2109

[23] R. Fielding, UC Irvine, J. Gettys, Compaq/W3C, J. Mogul, Compaq, H. Frystyk, W3C/MIT, L. Masinter, Xerox, P. Leach, Microsoft, T. Berners-Lee, W3C/MIT, « Hypertext Transfer Protocol -- HTTP/1.1 », Request for Comments 2616, Juin 1999
http://www.ietf.org/rfc/rfc2616

[24] D. Kristol, Bell Laboratories, Lucent Technologies, L. Montulli, Epinions.com, Inc., « HTTP State Management Mechanism », Request for Comments 2965, Octobre 2000
http://www.ietf.org/rfc/rfc2965

[25] Jigsaw Team, W3C (MIT, INRIA, Keio), « Jigsaw Documentation Overview » (last revised 23/09/1999)
http://jigsaw.w3.org/Doc/JigsawDoc.pdf

[26] Jigsaw Team, W3C (MIT, INRIA, Keio), « Jigsaw - W3C's Server », (Overview.html version 1.189 last revised 07/06/2002)
http://www.w3.org/Jigsaw/

[27] Privacy.net, « How Companies Can Track Your Movements on the Internet »
http://privacy.net/track/

[28] Privacy Foundation, « FAQ: Document Web Bugs »
http://www.privacyfoundation.org/resources/docbug.asp

[29] Electronic Privacy Information Center, « EPIC Online Guide to Pratical Privacy Tools » (last updated 30 july 2002)
http://www.epic.org/privacy/tools.html

[30] Vincent Bouthors, « Évolution du Web et de l'Internet », Novembre 1996
http://koala.ilog.fr/vb/WebSurvey-fr.html

[31] « The Web Bug FAQ »
http://personal.bellsouth.net/sdf/m/n/mnxian/webbugs.html

Le glossaire a été réalisé en se basant sur les sources suivantes :

[G1] e-SecurIT, « Glossaire sécurité »
http://securit.free.fr/glossaire.htm

[G2] Eric Goguey, Dictionnaire de l'Informatique
http://www.dicofr.com/

[G3] Liège et la société de l'information, Protection de la vie privée, Exposé de Mme Sophie LOUVEAU, chercheur au CRID, « Le traitement des données à caractère personnel sur le Net : Quelles obligations pour l'administration ? », Glossaire
http://www.forumtelecom.org/pv/010426bg.html

ANNEXE A : Glossaire des termes techniques

[A]

[ActiveX]
ActiveX est le nom donné par Microsoft à un ensemble stratégique de technologies et d'outils de développement orientés objet. La principale technologie est le Component Object Model (COM). Utilisé dans un réseau avec certaines extensions, COM devient Distributed Component Object Model (DCOM). L'élément principal que l'on crée en écrivant un programme s'exécutant dans l'environnement ActiveX est un composant, à savoir un programme autonome qui peut être exécuter n'importe où dans un réseau ActiveX (à l'heure actuelle, un réseau composé de systèmes Windows et Macintosh). Ce composant est connu sous le nom de contrôle ActiveX. ActiveX est la réponse de Microsoft à la technologie Java de Sun Microsystems. Un contrôle ActiveX est approximativement équivalent à une applet Java.

[ADSL]
ADSL (Asynchronous Digital Subscriber Line) est un protocole de communication qui peut être employé sur une ligne classique de paire torsadée en cuivre. Il permet d'atteindre une vitesse allant jusqu'à 1 Mbps, pendant que la ligne reste libre pour une communication téléphonique classique. ADSL nécessite des modems spécialisés ADSL à chaque extrémité de la ligne locale.

[Adresse IP]
Cette définition est basée sur le Internet Protocol version 4 (IPv4). Le Internet Protocol version 6 (IPv6) définit de nouvelles adresses IP d'une longueur de 128 bits.
Une adresse IP est un nombre binaire de 32 bits qui identifie chaque émetteur ou receveur d'informations au sein d'un réseau comme Internet. Elle comporte quatre nombres entre 0 et 255, séparés par des points (125.25.26.121 par exemple). L'adresse IP permet de localiser l'ordinateur et de s'y connecter. La plupart des adresses IP ont un nom textuel équivalent, appelé nom de domaine.
Ainsi, lorsque l'on demande une page HTML par exemple, le Protocole Internet inclut notre adresse IP dans la requête HTTP (en fait, dans chaque paquet s'il en faut plus qu'un pour transmettre la requête) et l'envoie à l'adresse IP obtenue en traduisant le nom de domaine situé dans l'URL indiquée. A l'autre bout, le receveur peut voir l'adresse IP de l'émetteur, qui demande la page, et peut lui répondre en lui envoyant la page voulue (cette opération nécessitant l'utilisation de l'adresse IP qu'il a reçue).
Une adresse IP est composé de deux parties :

- l'identificateur d'un sous-réseau particulier d'Internet,
- l'identificateur d'un ordinateur (un serveur ou une station de travail) appartenant à ce sous-réseau.

Lorsqu'un paquet IP voyage d'un routeur à un autre au travers d'Internet pour rejoindre sa destination, seul l'identificateur du sous-réseau a besoin d'être connu, l'identificateur de l'ordinateur destinataire étant uniquement nécessaire lorsque le paquet a atteint le sous-réseau dont cet ordinateur fait partie.

[Applet Java]
Une applet Java est une petite application utilisant Java (un langage de programmation orienté objet), qui peut être envoyée, au sein d'une page Web, à un internaute par un serveur ; cette application s'exécutera grâce à un interpréteur Java intégré au navigateur. Une applet Java peut exécuter des opérations extrêmement puissantes, des animations interactives, ou certaines tâches relativement simples sans devoir envoyer une requête en retour au serveur qui la transmise.

[Authentification]
L'authentification est la vérification de l'identité d'un utilisateur qui se connecte à un système informatique, ou bien celle de l'intégrité d'un message transmis.
C'est aussi un service de sécurité dont l'objectif est de valider l'identité d'une entité (utilisateur ou équipement). Il existe classiquement trois méthodes d'authentification permettant de prouver l'identité d'une entité :

- Authentification basée sur la connaissance d'un secret (ex. : mot de passe).
- Authentification basée sur la possession d'un objet (ex. : carte à puce).
- Authentification basée sur la biométrie.

[Autorisation]
L'autorisation est un service de sécurité visant à déterminer les droits d'une entité (utilisateur ou équipement) sur une ressource informatique (ex. : permissions sur un fichier). En général, ce service est lié au service d'authentification.

[Autorité de certification]
Une autorité de certification est une autorité chargée par un ou plusieurs utilisateurs de créer et d'attribuer des certificats. Composant décisionnel, l'autorité de certification joue les rôles suivants :

- Application de la politique de certification de l'organisme.
- Émission de certificats en associant l'identité du demandeur à une clé publique et en garantissant cette association par apposition de sa signature.
- Gestion du cycle de vie des certificats (attribution, durée de vie, révocation...).

[B]

[Backdoor]
Appelée porte dérobée en terminologie française, une backdoor est un programme malicieux visant à détourner les fonctionnalités d'un service ou d'un système en ouvrant des canaux d'accès masqués et utilisés par une personne malveillante.
Un backdoor est souvent mis en place à l'aide d'un cheval de Troie (ou Trojan).
BackOrifice et NetBus sont parmi les backdoors les plus célèbres.

[Bannière]
Les bannières publicitaires sont de petites fenêtres qui apparaissent au-dessus du contenu d'un site Web ou y sont intégrées.

[Base de données]
Une base de données est un ensemble structuré de fichiers informatiques contenant un ensemble cohérent de données. Hébergée par un serveur, elle peut contenir des quantités gigantesques de données. Par exemple, des sociétés de cybermarketing importantes peuvent conserver plus de 1000 Gb de données concernant les profils récoltés auprès de millions d'internautes.

[Bps (bits par seconde)]
Unité de mesure décrivant la vitesse à laquelle les données sont transmises par seconde. On mesure la vitesse d'un modem ou d'une liaison en kilobits/seconde (Kbps) ou mégabits/seconde (Mbps).

[C]

[Cache Web ou Antémémoire Web]
Un système informatique qui, au sein d'un réseau, conserve en mémoire ou sur disque une copie temporaire des dernières pages Web demandées dans le but d'accélérer leur importation. Si la page demandée a été stockée dans le répertoire cache du navigateur (parce qu'elle a été déjà visitée précédemment), elle est importée localement plutôt que du serveur d'origine sur Internet, d'où un gain de temps et un encombrement moindre du réseau. La taille du cache est généralement paramétrable selon les navigateurs.
Les serveurs cache Web sont appelés serveurs proxy et peuvent être situés à l'intérieur du firewall de l'entreprise. Ils permettent aux pages Web les plus couramment demandées d'être disponibles immédiatement. Comme le contenu des pages Web est sujet à modifications, le logiciel du cache vérifie en permanence s'il y a une nouvelle version pour la page, et la stocke si nécessaire.
Les pages seront effacées du cache après un certain temps d'inactivité.

[CERT]
Le Computer Emergency Response Team est une équipe de l'université de Carnegie-Mellon, créée en 1988 après la dangereuse diffusion d'un ver sur Internet. Elle est dédiée à la veille en sécurité informatique. Ainsi, CERT publie régulièrement des avis/alertes sur les failles de sécurité découvertes. Le modèle du CERT est classiquement repris au sein des grandes entreprises ou administrations pour constituer des équipes de veille en sécurité ou de réaction sur incident de sécurité (*Security incident response team*). En France, le CERT/A assure cette fonction vis-à-vis des grandes administrations françaises.
Site du CERT : http://www.cert.org/.

[Certificat électronique]
Un certificat électronique est un document électronique contenant deux types d'informations et destiné à prouver une identité dans le monde électronique. Il s'agit d'abord du certificat électronique lui-même, qui inclut le nom de la personne physique ou juridique demandant le certificat ou un pseudonyme, sa clé publique, les dates de validité du certificat et le nom de l'autorité certifiante (Certification Authority CA) l'ayant délivré. La deuxième partie est la signature électronique de l'autorité certifiante.
L'ensemble du message est électroniquement signé par l'autorité certifiante, qui jouit de la confiance de nombreux serveurs (l'autorité certifiante est une sorte de partie tierce de confiance) et peut vérifier la relation existante entre la personne physique ou juridique et sa clé publique.

[Client]
Un programme client désigne une application installée sur le poste de l'utilisateur qui utilise un service fournit par un serveur distant et/ou récupère des informations sur ce serveur.

[CGI (Common Gateway Interface)]
Langage de script exécuté sur le serveur qui permet de manipuler des données avant de les renvoyer à l'ordinateur client. C'est généralement un script CGI qui gère les données issues des formulaires que vous pouvez envoyer à partir d'une page Web.

[Chat]
Mode d'échange, en temps réel, de messages électroniques entre utilisateurs, où les interlocuteurs sont regroupés en canaux (« channel ») ; il s'agit d'une conversation collective ou privée en mode texte, interactive, par clavier interposé, sur un thème déterminé.

[Cheval de Troie]
Appelé Trojan en langue anglaise, un cheval de Troie est un programme d'aspect anodin, masquant un code exécutable malicieux déclenchant ou servant à déclencher une attaque.
Un cheval de Troie est en général utilisé pour ouvrir une porte dérobée (backdoor) sur un système.

[Chiffrement]
Action de chiffrer.

[Chiffrer]
Appliquer un code secret à un ensemble de données pour en assurer la confidentialité et l'authenticité.
Voir Chiffrage.

[Chiffrage]
Codage[1] des données (informations ou messages) de manière à ce qu'elles ne puissent en principe pas être facilement comprises par des personnes non autorisées, en l'occurrence une personne autre que le destinataire, qui dispose d'une clé ou d'un mot de passe pour procéder au décodage[2].
Il y a deux sortes de systèmes de cryptage principaux :

- Le système symétrique ou de clé privée, qui emploie une clé secrète connue de l'expéditeur et du destinataire d'un message; son principal avantage est la rapidité de traitement, et son principal inconvénient est la difficulté de communiquer les clés de manière sûre à un grand nombre d'utilisateurs. Dans le cas de SSL, 128 bits au maximum sont alloués à la clé secrète.
- Le système asymétrique ou de clé publique, qui emploie deux clés, générées de manière à ce que même en en connaissant une on ne puisse pas en pratique deviner l'autre. Les messages cryptés en employant l'une des clés sont décryptés en employant l'autre. L'une des clés est rendue publique et employée pour crypter les messages, que chaque utilisateur décrypte en utilisant sa propre clé privée secrète. La clé privée est également employée pour les signatures électroniques.

[Clickstreams]
Ces informations concernent un individu et reflètent son comportement, le chemin qu'il suit pour arriver sur un site ou ses choix lors de la visite d'un site. Elles contiennent le chemin des liens qu'un utilisateur a suivis dans chaque incursion à travers Internet, et sont stockées sur le serveur Web (l'ordinateur de l' ISP pour ceux qui n'ont pas leur propre serveur Web).

[Confidentialité]
Qualité de l'information qui n'est pas disponible ou accessible à des individus, entités ou processus non autorisés.

[1] Conversion de quelque chose (une donnée) dans une forme, appelée texte chiffré, rendue inintelligible.
[2] Processus consistant en la conversion d'un texte chiffré en sa forme originelle, afin de comprendre la donnée ayant été codée.

[Cookie]
Les cookies sont des informations générées et enregistrées par un site Web sur le disque dur (sous forme de fichiers texte) d'un client/visiteur, le plus souvent à son insu, pour en permettre la récupération lors de la prochaine visite du client, ce qui lui permet de s'en souvenir plus tard.
Utilisant le protocole HTTP, chaque requête pour l'obtention d'une page Web est indépendante de toutes les autres requêtes. Pour cette raison, les serveurs Web n'ont pas la « mémoire » des pages qu'ils ont envoyées à un internaute précédemment, ni de n'importe quelle autre information concernant ses visites antérieures. Ainsi, les cookies offrent un mécanisme permettant à un site Web de stocker ses propres informations à propos d'un visiteur sur le propre ordinateur de ce dernier.
Comme les cookies font partie intégrante du trafic HTTP, ils peuvent donc être transportés sans problème avec le trafic IP. En particulier, les cookies peuvent s'avérer très utiles, notamment pour maintenir une session utilisateur sur un flux HTTP, pour mémoriser un code secret ou pour donner accès à un contenu personnalisé. Inversement, couplés à des données nominatives, les cookies constituent une réelle menace pour la protection de la vie privée En effet, un cookie peut contenir un numéro identificateur unique (GUI, Global Unique Identifier) permettant un meilleur pistage que les adresses IP dynamiques. Il donne ainsi à un site Web le moyen de garder trace des comportements et préférences d'un utilisateur.
On peut voir les cookies qui ont été enregistrés sur son disque dur (même si leur contenu peut n'avoir aucune signification intelligible). L'endroit de leur stockage dépend du navigateur utilisé. Ainsi Internet Explorer enregistre chaque cookie dans un fichier séparé à l'intérieur d'un sous-répertoire spécifique, tandis que Netscape les stocke tous dans un seul et unique fichier.
Les cookies contiennent une série d'URLs (adresses), pour lesquelles ils sont valables. Lorsque le navigateur rencontre à nouveau ces URLs, il envoie ces cookies spécifiques au serveur Web.
Les cookies peuvent être de nature différente : ils peuvent être persistants (et avoir une durée de vie de plusieurs dizaines d'années), mais peuvent également avoir une durée limitée et dans ce cas, on les nomme cookies de session.
On peut demander à son navigateur de refuser les cookies ou de nous prévenir avant d'en accepter un. Les cookies peuvent également être supprimés manuellement.

[Coupe-feu (pare-feu ou Firewall]
Un coupe-feu est un système configuré spécialement pour contrôler le trafic circulant entre plusieurs réseaux (le plus souvent entre un réseau d'entreprise et Internet).
C'est une méthode de sécurisation d'un réseau. Son rôle est de filtrer les paquets indésirables et de bloquer les attaques malveillantes. Il peut être mis en oeuvre par un serveur, un routeur ou une combinaison de routeurs et d'hôtes. Les coupe-feu sont le plus souvent utilisés pour donner aux utilisateurs un accès sécurisé et protégé à Internet[3], ainsi que pour séparer le serveur Web public d'une société de son réseau local. Ils sont également employés pour sécuriser ou isoler des segments du réseau interne.
Il existe aussi un type particulier de firewalls, les firewalls personnels, existant sous forme d'un service logiciel installé directement sur l'ordinateur des utilisateurs et filtrant les connexions issues de leur ordinateur à destination de n'importe quel autre ordinateur.

[Courrier électronique (e-mail)]
Courrier acheminé sur un réseau informatique. Un courrier électronique peut contenir du texte, des images, du son, du code HTML, etc. Il peut également être accompagné en annexe (« attachment ») d'un ou plusieurs fichiers.
Le courrier électronique a été l'une des premières utilisations d'Internet. En attribuant des adresses électroniques personnelles aux internautes, Internet peut ainsi transmettre des messages d'un internaute à l'autre. Ce processus est habituellement très rapide étant donné que le courrier électronique est souvent composé uniquement de texte en ASCII, ce qui n'exige pas beaucoup de ressources.

[3] Ainsi, les utilisateurs d'un réseau local d'entreprise ont accès aux ressources d'Internet en passant par le firewall, mais ce dernier empêche les internautes d'entrer dans le réseau sans autorisation.

[Cryptanalyse]
La cryptanalyse est l'étude de la sécurité des procédés cryptographiques. La cryptanalyse consiste à déchiffrer un message dont on connaît généralement le procédé de chiffrement, mais pas les secrets.

[Cryptographie]
La cryptographie est une discipline qui englobe tous principes, moyens et méthodes destinés à la transformation de données afin de cacher leur contenu, d'empêcher leur modification et leur utilisation frauduleuse. C'est une science qui utilise les mathématiques pour effectuer les opérations de chiffrage et de déchiffrage, c'est-à-dire pour dissimuler un texte en clair en rendant son contenu inintelligible.

[Cryptographie symétrique]
Voir Technique de cryptographie symétrique.

[Cryptologie]
Étude des procédés de chiffrement. Ensemble de la cryptanalyse et de la cryptographie.

[Crypto-système à clé publique (crypto-système asymétrique)]
Un crypto-système à clé publique est un système cryptographique consistant en deux opérations complémentaires, chacune utilisant l'une des deux clés distinctes mais associées, la clé publique et la clé privée et possédant la propriété selon laquelle il est impossible de déterminer, par un calcul sur ordinateur, la clé privée à partir de la clé publique.

[Cyberwoozle/Cyberwoozling]
Cyberwoozle est un terme commercial développé par Content Technologies désignant la collecte illicite d'informations stockées sur un poste de travail, effectuée lorsque celui-ci se connecte à un site Web.
Les moyens utilisés sont en général basés sur les cookies, les plug-ins des navigateurs...

[D]

[Décrypter]
Action consistant à retrouver un ensemble de données en clair à partir d'un message chiffré, sans connaître le code secret de chiffrement.

[Déni de service]
Le déni de service (Denial of Service (DoS), en anglais) est une attaque consistant à saturer une ressource en effectuant de manière malveillante des demandes de réservation excessives ou en occupant le service illicitement.
Parmi les attaques de déni de service les plus connues : ping of death, ...

[Données à caractère personnel]
Toute information concernant une personne physique identifiée ou identifiable (le n° de téléphone d'une personne, sa fonction, la mention de sa présence et de ses interventions à une réunion, sa photo, etc.).

[DNSSEC]
Domain Name Service Security Extensions est un service de résolution de noms de domaine, assurant la traduction des adresses IP en noms et inversement. Dans le fonctionnement d'Internet, il s'agit d'un élément essentiel et critique. Ainsi, DNS est sensible aux attaques d'usurpation d'identité (*DNS spoofing*) consistant à rediriger un nom de machine vers l'adresse IP d'une machine pirate.
D'où la création du DNSSEC, standardisé par l'IETF (RFC 2065), qui utilise la cryptographie asymétrique pour assurer :

- l'authentification de l'origine des données.
- l'intégrité des données.

DNSSEC s'appuie sur la technique de signature numérique pour garantir l'origine des informations.

[F]

[Fichier de log (Fichier d'Audit, Log ou Log file)]
Fichier contenant les informations de connexion de tous les visiteurs d'un site Web. Il est généré par le serveur hébergeant le site et présente l'ensemble des informations relatives à ce site, notamment pour sa fréquentation (hits, pages vues, etc.)

[Firewall]
Voir Coupe-feu.

[Fournisseur de services Internet (Internet Service Provider - ISP)]
Une société qui fournit un accès et une connexion à Internet à des particuliers et à des sociétés contre une rémunération mensuelle. Le service fourni comprend souvent, outre l'accès à Internet proprement dit, une adresse e-mail et un mot de passe pour envoyer et recevoir du courrier électronique, des accès à certains serveurs de jeux ou de téléchargement de logiciels ou autre.
Les petits fournisseurs de services Internet fournissent un service via un modem, une ligne ISDN ou ADSL, alors que les plus grands proposent également des lignes de connexion privées. Les utilisateurs reçoivent généralement une facture fixe mensuelle, mais d'autres coûts peuvent être impliqués. Contre paiement, un site Web peut être créé et hébergé sur le serveur de l'ISP, permettant ainsi aux petites sociétés d'être présentes sur le Web avec leur propre nom de domaine.
Les grands fournisseurs de services Internet fournissent également des bases de données privées, des forums et des services supplémentaires, en plus de l'accès à Internet.

[FTP (File Transfer Protocol)]
FTP est le protocole standard de transfert de fichiers, il est la manière la plus simple d'échanger des fichiers entre ordinateur sur Internet. C'est donc un ensemble de conventions qui définissent les règles permettant à un serveur FTP de dialoguer avec un client FTP (les navigateurs récents sont tous clients FTP en même temps que clients HTTP). Un serveur FTP permet à un client d'écrire ou de lire des fichiers sur le (ou à partir du) serveur. Comme le HTTP, qui transfère les pages Web et les fichiers reliés, ou le SMTP, qui expédie les e-mails, FTP est un protocole applicatif qui repose sur les protocoles TCP/IP pour les échanges de données sur le réseau. FTP est souvent utilisé par les webmasters pour transférer les fichiers de leur site (pages Web, images,...) vers le serveur qui l'hébergera, mais aussi par les internautes pour télécharger des programmes et d'autres fichiers à partir d'un serveur vers leur PC.
Pratiquement, FTP permet à l'internaute de visiter un répertoire situé sur un serveur et, selon les droits dont il dispose, y déposer ou prendre des documents. Un grand nombre de sites FTP sont dits "anonymes" parce qu'ils permettent aux utilisateurs d'y accéder en lecture sans identification préalable.

[H]

[Hacker]
En argot informatique, nom donné aux passionnés d'Internet et d'informatique qui visitent et/ou piratent les banques de données ou les sites Web du monde entier et cherchent les failles de sécurité dans les systèmes informatiques. Contrairement aux crackers, les hackers n'agissent pas toujours avec l'intention de nuire.

[Hit (coup)]
Dans le contexte de la fréquentation d'un site Web, un hit est une simple demande d'accès à un fichier texte, graphique ou autre du serveur. Chaque fois qu'un navigateur charge une page Web, il génère un hit pour cette page et autant de hits qu'il y a d'objets sur cette page (par ex : si une page Web contient 8 images séparées, un accès à cette page représentera 9 hits sur le serveur : 8 images + la page). Le hit ne doit pas être confondu avec le nombre de pages vues qui sert en général de référence pour la fréquentation des sites Web.

[HTML (HyperText Markup Language)]
HTML est le langage de description de pages dont les deux caractéristiques principales sont
1/ la navigation hypertextuelle grâce aux hyperliens,
2/ le multimédia (intégration de textes, d'images, et de sons).
Le navigateur lit la page HTML, interprète les balises et les liens, et affiche la page à l'écran.
(Plus techniquement, le HTML est un langage de balisage de texte qui permet la création de documents hypertextes enrichis affichables par un navigateur Web.) Ce langage est composé de tags (ou balises) insérés dans un fichier et qui permettent d'en décrire le contenu. Ce sont ces tags qui indiquent au navigateur comment il doit afficher le texte et les images constituant une page Web mais aussi où se trouvent les hyperliens et vers quelles URLs ceux-ci pointent. Des balises HTML permettent également de coder des méta-données, qui seront utilisées par les moteurs de recherche par exemple pour répertorier les pages dans leur index, grâce aux « méta tags ».
HTML est compréhensible par tous les navigateurs, néanmoins certains d'entre eux supportent certaines extensions additionnelles non standard ou implémentent différemment certaines caractéristiques. La version actuelle du HTML est HTML 4.0 et est une recommandation « officielle » du World Wide Web Consortium.

[HTTP (HyperText Transfert Protocol)]
HTTP est un protocole applicatif grâce auquel un client Web et un serveur Web peuvent dialoguer. Il est constitué d'un ensemble de règles pour l'échange de fichiers (texte, images, sons, vidéos, et autres fichiers multimédia) sur Internet. Il repose sur les protocoles TCP/IP, qui sont la base de toute transmission d'informations sur un réseau. HTTP utilise le port 80 dans ses interactions avec la couche inférieure, TCP/IP.

[HTTPS (Secure HyperText Transfer Protocol)]
HTTPS est une version sécurisée de HTTP assurant les services de sécurité :
- Authentification (éventuellement mutuelle).
- Confidentialité (chiffrement des données échangées).
- Intégrité des données (au cours de leur transport).

HTTPS a été développé par Netscape et intégré pour la première fois dans leur navigateur, celui-ci chiffre et déchiffre les requêtes d'un utilisateur ainsi que les données qui lui sont retournées par un serveur Web. HTTPS s'appuie sur le protocole SSL (techniquement, HTTPS c'est l'utilisation de SSL comme une sous-couche du protocole applicatif HTTP) et les algorithmes cryptographiques associés, de sorte qu'il supporte certains types de certificats. HTTPS utilise le port standard 443.

[Hyperlien]
Un hyperlien est une liaison prédéfinie entre un objet et un autre (le plus souvent entre deux documents sur le Web). Le lien est affiché sous forme de texte, d'image ou d'icône. Sur les pages Web, un texte hyperlien s'affiche généralement comme un texte en bleu souligné, alors qu'un hyperlien graphique se présente comme une petite image. Il peut être activé d'un simple clic sur l'objet désigné.
Cette forme de navigation facilite la circulation à l'intérieur d'un site Web ou entre différentes pages situées sur des sites différents.
Un hyperlien invisible est un hyperlien caché dans le code HTML d'une page Web (invisible pour l'internaute) et indiquant le plus souvent une image située sur un serveur autre que celui hébergeant le site visité, image que le navigateur téléchargera automatiquement (et sans en avertir l'internaute).

[I]

[ICQ]
ICQ ("I seek you") est un programme qui permet de savoir quand ses contacts (amis, famille, collègue,...) sont aussi en ligne sur Internet, de les appeler, et de discuter avec eux. Il permet d'envoyer des messages simples, de petits fichiers, et des URLs directement vers le contact choisi. De plus, on peut initier une session de discussion avec connexion vocale ou visuelle/vocale ou jouer à des jeux avec d'autres membres de la communauté ICQ. Les événements entrants sont signalés dès leur arrivée et la connexion peut être établie immédiatement. Pour utiliser ce logiciel, il faut que les deux interlocuteurs l'aient installés sur leur ordinateur et se soit enregistrés afin de posséder un numéro d'identification unique.

[IETF (Internet Engineering Task Force)]
Groupe responsable de l'établissement des normes et des standards d'Internet, il doit également proposer des solutions aux problèmes liés à Internet, assurer le transfert technologique vers la communauté d'Internet en général et de favoriser les échanges entre les différents acteurs d'Internet.

[IP]
Internet Protocol est le protocole par lequel des données sont envoyées d'un ordinateur à un autre au moyen d'Internet ou d'un réseau du même genre. Chaque ordinateur connecté à Internet (appelé hôte) a au moins une adresse IP qui l'identifie univoquement parmi tous les autres ordinateurs sur Internet. Quand on envoie ou reçoit des données, le message est divisé en petits morceaux appelés paquets. Chacun de ces paquets contient à la fois l'adresse IP de l'émetteur et du destinataire. Un paquet est d'abord envoyé à un routeur qui lira l'adresse de sa destination et le transfèrera à un routeur adjacent, qui à son tour lira l'adresse de destination, et ainsi de suite, le paquet voyagera à travers le réseau jusqu'à atteindre un routeur reconnaissant la destination du paquet comme un ordinateur appartenant à son domaine ou à son voisinage immédiat. Ce routeur enverra alors le paquet directement à l'ordinateur dont l'adresse IP est celle de la destination du paquet.

[Infomédiaire]
Les infomédiaires sont des sociétés opérant sur Internet qui protègent gratuitement la vie privée de leurs clients. Pour générer des revenus, les infomédiaires commercialisent des profils, sans données nominatives, aux professionnels du marketing. Le concept a été forgé par John Hagel III et Marc Singer dans leur ouvrage "Net Worth".

[Infrastructure de gestion de clés]
Une infrastructure de gestion de clés offre un environnement de confiance, ainsi qu'un ensemble de garanties et services relatifs aux certificats électroniques.
Une infrastructure de gestion de clés est composée des éléments suivants :

- Autorité de certification.
- Système de publication/distribution des certificats (ex. annuaire).
- Applications compatibles.

Une infrastructure de gestion de clés utilise les objets suivants :
- Paires de clés.
- Certificats.

[Intégrité de données]
Qualité de données qui n'ont pas été altérées ou détruites de manière frauduleuse.
Intégrité des données. C'est également un processus destiné à éviter la destruction ou l'altération accidentelle d'une base de données

[Internet]
Ensemble hétérogène de réseaux de toutes tailles (il s'agit de réseaux nationaux, régionaux, publics et privés) interconnectés entre eux respectant le protocole d'adressage IP et capable de communiquer à l'aide du protocole de communication TCP.
Les services d'Internet sont, notamment : l'e-mail, les newsgroups, telnet, FTP et Web.

Autre définition : Réseau mondial associant des ressources de télécommunication et des ordinateurs serveurs et clients, destiné à l'échange de messages électroniques, d'informations multimédias et de fichiers. Il fonctionne en utilisant un protocole commun qui permet l'acheminement de proche en proche de messages découpés en paquets indépendants.
L'acheminement est fondé sur le protocole IP (Internet Protocol), spécifié par l'Internet Society (ISOC).
L'accès au réseau est ouvert à tout utilisateur ayant obtenu une adresse auprès d'un organisme accrédité.
La gestion est décentralisée en réseaux interconnectés.

[IPSEC]
Internet Protocol SECurity est un ensemble de protocoles normalisés sous la conduite de l'IETF, afin d'améliorer la sécurité du protocole IPv4 (natif en IPv6) face aux attaques de type écoute (IP-sniffing), usurpation d'identité (IP-spoofing), prédiction de séquences de paquets, re-jeu de trafic...
IPSEC garantit l'authenticité, l'intégrité, la confidentialité et le non re-jeu des paquets IP échangés de bout en bout entre deux entités en s'appuyant sur les techniques de cryptographie asymétrique.
IPSEC définit principalement :

- Deux protocoles d'encapsulation sécurisée : AH (Authentication Header) et ESP (Encryption Security Payload).
- Deux structures de gestion de la sécurité par les protagonistes de la communication IPSEC : SA (Security Association) et SPD (Security Policy Database).
- Des procédures d'échange et de gestion des clés : IKE (Internet Key Exchange).

IRC]
Internet Relay Chat est un système de discussion (ou « chat », en anglais) qui inclut un ensemble de règles et de conventions, et nécessite l'emploi d'un logiciel client/serveur. Sur Internet, certains sites ou réseaux IRC fournissent des serveurs et permettent de télécharger le logiciel nécessaire (un client IRC).

[ISDN]
Integrated Services Digital Network est un ensemble de standard CCITT/ITU pour la transmission digitale d'informations sur une ligne téléphonique ordinaire mais aussi sur d'autres medias. Les internautes qui ont installé un adaptateur ISDN (à la place d'un modem) bénéficient d'un débit d'échanges de données pouvant atteindre 128Kbps.

[J]

[Java]
Java est un langage de programmation de haut niveau, complet et basé sur le modèle de programmation orienté objet, créé par la société Sun Microsystems. Ce langage est conçu spécifiquement pour une utilisation dans un environnement distribué. Les programmes Java offrent l'avantage de pouvoir être exécutés sur n'importe quel type d'ordinateur grâce à une machine virtuelle (langage portable). Sa conception lui donne le « look and feel » du langage C++ tout en étant plus simple à utiliser que le C++. Java peut être utilisé pour développer des applications complètes qui peuvent s'exécuter sur un seul ordinateur ou être distribuées parmi les serveurs et les clients dans un réseau. Il peut également servir à construire un petit module d'application (une Applet) pour un usage au sein d'une page Web.

[Javascript]
JavaScript est un langage de script utilisant une syntaxe similaire à celle de Java, mais qui n'est pas compilé. Il reste sous forme de code source au sein d'un document HTML et doit donc être interprété ligne par ligne par l'interpréteur JavaScript incorporé au navigateur.
JavaScript est très populaire et est accepté par la majorité des navigateurs Web. JavaScript est plus limité que Java et s'applique essentiellement aux éléments de la page Web elle-même. Le langage Javascript est simple mais peu puissant. Il peut être utilisé dans différents buts au sein d'un site Web tels que : inclure un compteur des visites, changer automatiquement une date, ouvrir une nouvelle fenêtre vers une URL donnée (ce qu'on appelle une fenêtre pop-up), modifier un texte ou une image lorsque le curseur de la souris passe dessus (effet rollover),..., mais aussi obtenir des informations sur le navigateur d'un visiteur ou encore créer, modifier ou lire des cookies.

[L]

[Label Internet]
Un label Internet est un logo figurant sur un site Web, en général sur la page d'accueil, symbolisant l'adhésion du site à un code de bonne conduite élaboré par une organisation extérieure. Le label a pour vocation de contribuer au sentiment de confiance de l'Internaute.

[M]

[Meta Tags]
Les meta tags sont des balises HTML qui donnent des informations sur une page Web.
Contrairement aux tags HTML normaux, les meta tags n'ont pas d'incidence sur la manière dont la page est affichée. En revanche ils donnent des informations sur le créateur de la page, le nombre de mises à jour, son sujet, et les mots clés caractérisant le contenu de la page. De nombreux moteurs de recherche emploient ces informations pour construire leur index.

[MIME]
Multi-Purpose Internet Mail Extensions est une extension du protocole d'e-mail originel (Simple Mail Transfer Protocol) qui permet d'utiliser ce protocole pour échanger différents types de fichiers de données sur Internet (audio, vidéo, images, programmes, etc., aussi bien que des textes formatés en ASCII tel que supporté à l'origine par le SMTP).

[Modem]
Le MOdulateur-DEModulateur est un appareil qui fait la jonction entre un terminal ou un ordinateur, et une ligne téléphonique analogique en convertissant des impulsions électroniques en fréquences audio, et vice-versa. Ce terme fait fréquemment référence à des modems de 56 Kbps (V.90), qui sont les appareils actuels de haut niveau, ou à des modems plus anciens de 28,8 Kbps (V.34). Le terme peut également s'appliquer à des modems encore plus rapides pour le câble de télédistribution, l'ADSL ou à des adaptateurs terminaux pour lignes ISDN, qui sont entièrement électroniques, et ne sont donc techniquement pas des modems. Un modem est un adaptateur analogique-électronique et électronique-analogique. Il peut également appeler une ligne, répondre à un appel et contrôler la vitesse de transmission des données.
Quelle que soit la vitesse maximale, le modem accepte aussi un certain nombre de vitesses plus basses pour pouvoir être en contact avec des modems plus anciens, ou s'adapter à une vitesse inférieure pour des lignes téléphoniques de moindre qualité.

[Mot de passe]
Un mot de passe est un moyen utilisé pour authentifier une entité. Il s'agit d'une suite secrète de caractères.

[Moteur de recherche]
Outil de recherche dont l'interrogation donne une page HTML, laquelle est affichée par le navigateur de celui qui formule la question. Cette page de résultats est la liste des pages qui répondent aux critères d'interrogation qui ont été définis initialement. L'interrogation consiste en fait à demander à l'outil de recenser les pages qu'il a indexées et qui contiennent la ou les séquences de caractères qui lui ont été spécifiées: termes, fractions de termes ou expressions complètes, tous ces éléments pouvant être ou non placés entre parenthèses ou, encore, associés d'une manière ou d'une autre. Dans le contexte d'Internet, les moteurs de recherche assument une fonction primordiale pour trouver une information dans l'immensité représentée par la multitude de sites Web.

[N]

[NAT]
Network Address Translation.
Voir traduction d'adresse.

[Navigateur (ou browser)]
Logiciel client conçu pour permettre aux utilisateurs de naviguer de manière simple au sein du Web, de rechercher des ressources Web et d'y accéder. Le navigateur permet donc à l'internaute de passer d'une page Web à l'autre en cliquant sur un lien. Si l'exploration du Web passe obligatoirement par les navigateurs, il ne faut pas perdre de vue qu'ils se perfectionnent et rendent aujourd'hui d'autres services: téléchargement de fichiers par FTP, courrier électronique, forums de discussion. Les navigateurs les plus connus actuellement sont Internet Explorer, Netscape (Communicator), Opera et Mozilla.

[NNTP]
Network News Transfer Protocol est le protocole principal utilisé par les ordinateurs pour manipuler les notes postées sur un newsgroup (ou forum de discussion). Les serveurs NNTP collectent et contrôlent l'ensemble des messages postés sur un newsgroup. Un client NNTP est nécessaire pour accéder aux forums de discussion, la plupart des navigateurs en incluent un, mais il existe également des logiciels spécifiques.

[Non-répudiation]
Service de sécurité dont l'objectif est de générer, récolter, maintenir, rendre disponible et valider l'évidence (information utilisée pour établir une preuve) concernant un évènement ou une action revendiquée afin de résoudre les possibles disputes sur l'occurrence ou non de l'évènement ou de l'action.

[O]

[One-to-One]
Le One-to-One, c'est le marketing sur mesure. C'est aussi une stratégie qui consiste à s'adresser à chaque client en fonction de ses caractéristiques personnelles par opposition au "mass-marketing" qui était l'un des fondements de l'économie industrielle. Le véritable "one -to-one" très présent au niveau du discours reste en pratique un concept assez peu mis en oeuvre à la différence du "one-to-few" fondé sur les techniques de segmentation.

[Opt-in]
Garantie de consentement.
Système où l'accord préalable des personnes est nécessaire avant l'inscription dans une base de données et la mise en oeuvre d'actions de marketing direct.
La collecte d'une adresse électronique et l'utilisation commerciale qui peut en découler sont subordonnées au consentement préalable de la personne concernée.

[Opt-out]
Droit d'opposition.
Système où l'accord des personnes est supposé acquis par principe, sauf avis contraire, pour l'inscription dans une base de données et la mise en oeuvre d'actions de marketing direct.
Toute adresse électronique accessible sur Internet peut être utilisée à des fins de prospection dès lors que les personnes concernées se trouvent dans la possibilité de s'opposer à la réception de tels messages. Cet objectif peut être atteint par la mise en place de registres spécifiques régulièrement mis à jour.

[P]

[P3P]
Platform for Privacy Preference. Dénomination d'un projet mené par the World Wide Web Consortium, qui aspire à devenir le standard minimal de l'Internet en matière de protection de la vie privée. Cette plateforme permet aux internautes d'être informés des pratiques d'un site en matière de vie privée, de déléguer leurs décisions à leur navigateur (quand c'est possible) et d'adapter en conséquence leurs rapports à des sites particuliers.

[Paquet]
Unité d'information utilisée pour communiquer sur le réseau. Les messages émis entre les périphériques du réseau (postes de travail ou serveurs) forment des paquets sur le périphérique source. Les paquets sont assemblés, si nécessaire, pour former des messages complets lorsqu'ils atteignent leur destination. Un paquet peut contenir une requête de service, des informations sur le mode de traitement de la requête et les données qui vont faire l'objet du service. Un paquet est constitué d'en-têtes[4] et d'une portion réservée aux données. Différents en-têtes sont ajoutés à la portion de données au fur et à mesure que le paquet traverse les différentes couches de communication. Un message qui excède la taille maximale fixée est divisé et transmis en plusieurs paquets. Lorsqu'un paquet arrive à destination, les en-têtes sont supprimées dans l'ordre inverse de celui où ils avaient été ajoutés et la requête est traitée.

[Page vue]
Le nombre de pages vues est l'unité de mesure la plus fréquemment utilisée pour mesurer la fréquentation d'un site. Pour compter une page vue, il faut que l'ensemble des informations qui y sont contenues ait été complètement chargé par le browser de l'internaute

[Paire de clés]
Une paire de clés est l'ensemble composé d'une clé publique et de sa clé privée créées par et utilisées avec un système cryptographique asymétrique.
On distingue en général plusieurs natures de paires de clés :

- Paires de clés de confidentialité, utilisés pour chiffrer des messages de petite taille.
- Paires de clés de signature, dont la clé privée est utilisée pour signer les messages et la clé publique pour vérifier les signatures.
- Paires de clés de certification, utilisés par l'autorité de certification pour signer des certificats ou des messages de révocation.
- Paires de clés d'échange/transport de clés, utilisés pour le transport de clés symétriques servant à sécuriser les communications.

[4] Partie d'un paquet qui précède les données que vous envoyez. L'en-tête contient des informations sur l'expéditeur, le destinataire, et des instructions pour le contrôle des erreurs.

[Parties Tierces de Confiance]
Une partie tierce de confiance PTC (Trusted Third Party TTP) peut être décrite comme une entité en laquelle d'autres entités ont confiance en ce qui concerne des services et activités liés à la sécurité.
Une PTC offrira des services de grande valeur à des utilisateurs souhaitant améliorer le niveau de sécurité et de confiance des services qu'ils reçoivent, et sécuriser leurs communications avec leurs partenaires en affaires. Les PTC doivent pouvoir prouver un haut niveau d'intégrité, de respect de la confidentialité, et de garantie en ce qui concerne les services et informations impliqués dans les communications entre applications commerciales. En outre, les utilisateurs requerront les services des PTC lorsqu'ils en auront besoin, dans le cadre d'un contrat de service convenu.
Une PTC sera d'habitude une organisation sous licence de, ou accréditée par une autorité réglementaire, qui offrira des services de sécurité, sur base contractuelle, à une large gamme d'organismes, notamment des organismes dans le secteur des télécommunications, de la finance et du commerce.
Une PTC pourra par exemple être utilisée pour assurer l'attribution de signatures électroniques, de façon à garantir l'intégrité des documents.
De plus, elles peuvent fournir aux utilisateurs des services de cryptage dans les deux sens, et proposer des fonctions de dépôt/recouvrement de clé pour permettre de récupérer des fichiers en cas de perte de la clé de cryptage (typiquement des documents et fichiers cryptés par des employés), ou pour aider un utilisateur dans le cas d'une demande d'interception légale.
Le fait de faire appel à une PTC dépend fondamentalement de la confiance que lui portent les organismes auxquels elle rend des services.

[PING]
Ping est un programme de base qui permet de vérifier qu'une adresse IP particulière existe et peut accepter des requêtes (càd qu'un ordinateur possédant cette adresse IP est connecté à Internet et répond aux requêtes). C'est donc un utilitaire de diagnostic servant à s'assurer que l'ordinateur hôte que l'on essaie de contacter est effectivement en activité ou, dans le cas d'un hôte en activité, pour voir combien de temps il faut pour obtenir une réponse à une requête qu'on lui a envoyé (tester son temps de réponse).

[Plugins]
Des applications plug-in sont des programmes qui peuvent être facilement installés et utilisés comme une partie du navigateur. Généralement, elles permettent d'ajouter des fonctionnalités supplémentaires (le plus souvent il s'agit de fonctions multimédia) à un navigateur qui en est dépourvu (par exemple, le Flash Player permet de visionner et d'interagir avec des animations Flash dans une page Web). Une application plug-in est reconnue automatiquement par le navigateur dans lequel elle est intégrée, celui-ci fait alors directement appel à elle lorsqu'un contenu particulier situé dans une page HTML nécessite ses fonctions pour être présenté.

[Portail]
Un site portail donne une vue d'ensemble de liens Web d'une manière ordonnée. Via le portail visité, l'internaute peut aisément visiter des sites Web sélectionnés d'autres fournisseurs de contenu.
Les portails modernes sont des « supersites » proposant toute une gamme de services, notamment la recherche sur le Web, des nouvelles, les annuaires pages blanches et jaunes, des e-mails gratuits, des groupes de discussion, du shopping en ligne et des liens vers d'autres sites.
Sur le World Wide Web, un portail est un site proposant d'être le point de départ d'un grand nombre d'internautes, idéalement dès leur connexion. Après les portails généralistes, on a assisté à la naissance de portails spécialisés dans des secteurs précis (sport, shopping, santé, etc.).

[Privacy – ePrivacy]
La traduction littérale du terme en français serait « intimité » mais le concept général de « protection de la vie privée » semble plus approprié. On parle parfois de e-Privacy lorsqu'il s'agit de la protection des données personnelles et de la vie privée sur Internet.

[Privacy Policy]
Politique relative à la vie privée
Politique d'une entreprise en matière de protection des données personnelles. C'est également une charte déterminant la politique mise en place pour un responsable de site, un hébergeur, etc., afin de protéger la vie privée des visiteurs et utilisateurs.

[Pistage]
Traque des parcours des internautes sur Internet (sites visités,…).

[Profilage]
Caractérise l'internaute par la collecte d'informations le concernant afin de déterminer ses centres d'intérêts, ses capacités financières et/ou n'importe quelles caractéristiques personnelles.

[Porte-dérobée]
Voir Backdoor.

[PPP]
PPP (Point-to-Point protocol) est un protocole de télécommunication utilisé pour connecter deux ordinateurs par leur port série ou par un modem qui y est installé. C'est le protocole de couche inférieure employé entre le PC d'un l'utilisateur privé et le fournisseur d'accès Internet (ISP) lorsque la connexion est établie sur des lignes téléphoniques classiques. Par exemple, un ISP fournit une connexion PPP à un internaute pour que son serveur renvoie ses requêtes vers Internet et lui transmette en retour les réponses demandées. Selon le modèle OSI, PPP fournit un service concernant une couche inférieure à TCP/IP. Par conséquent, il emballe les paquets TCP/IP émis et les envoie sur la ligne téléphonique au serveur qui se trouve à l'autre bout de la connexion.

[Protocole]
Dans le présent contexte, un protocole est une série de règles techniques qui doivent être respectées par les deux partenaires lors d'un échange d'informations. Les protocoles sont organisés en une hiérarchie de couches, ainsi qu'on les nomme. Chaque couche est responsable du traitement d'un aspect particulier du processus de télécommunication et fournit des fonctionnalités qui sont utilisées par les couches supérieures. Sur Internet, la couche intermédiaire est classiquement celle du protocole TCP/IP.
Ainsi, Ethernet (pour les réseaux locaux LAN), ADSL (pour les lignes téléphoniques), ATM (pour les opérateurs de télécom), X-75 (pour les lignes ISDN) et PPP (pour les lignes téléphoniques classiques) sont des exemples de protocoles employés dans la couche inférieure.
De l'autre côté, HTTP (pour surfer), SMTP et POP (pour les e-mails), FTP (pour les transferts de fichiers) sont des protocoles de la couche supérieure. Ceci signifie que toute menace pour le respect de la vie privée présent dans le protocole TCP/IP sera une faiblesse pour les protocoles des couches supérieures. Concrètement, les couches sont des séries de sous-programmes qui tournent sur un ordinateur connecté à Internet.

[Protocole de configuration automatique d'hôte (Dynamic Host Configuration Protocol DHCP)]
Le Protocole de configuration automatique d'hôte (DHCP) est un protocole Internet destiné à configurer automatiquement les ordinateurs utilisant TCP/IP. DHCP assigne automatiquement des adresses IP.

[R]

[Réseau TCP/IP]
Un réseau TCP/IP (Transmission Control Protocol/Internet Protocol) est basé sur la transmission de petits paquets d'informations. Chaque paquet comporte l'adresse IP de l'expéditeur et du destinataire. Ce réseau n'est pas connecté. Cela signifie que, contrairement à un réseau téléphonique par exemple, il n'est pas nécessaire d'établir au préalable une connexion entre deux appareils avant d'entamer une communication. Cela signifie également que plusieurs communications sont possibles en même temps avec plusieurs partenaires.

[Request for Comments (RFC)]
Document public informatif ou descriptif détaillant les standards développés ou en développement, les comptes-rendus des réunions, les opinions,…

[Routeur]
Un routeur est un appareil important chargé de définir des routes pour les réseaux TCP/IP. Ceci implique que la route du TCP/IP est dynamique, et peut dépendre d'un échec de connexion ou d'une surcharge de certains routeurs ou liens. Il peut également servir de coupe-feu entre une organisation et Internet, et garantir que seules des adresses IP autorisées peuvent provenir d'un ISP en particulier.

[RSA]
Algorithme créé par Rivest, Shamir et Adleman (RSA) en 1977, et basé sur la cryptographie asymétrique, RSA est l'algorithme le plus utilisé dans le monde et offre les services de sécurité essentiels tels que l'authentification, la confidentialité, l'intégrité et la signature.
RSA connaît un essor important avec le développement des infrastructures de gestion de clés (PKI). Anciennement propriété de la société RSA Security, mais ouvert au domaine public depuis la fin 2000, l'algorithme RSA repose sur la difficulté de factoriser un grand nombre suivant ses facteurs premiers. Les clés asymétriques utilisées dans le cadre de RSA sont classiquement de taille 512, 1024, 2048, 4096 bits. On considère à l'heure actuelle qu'une clé de 512 bits à une durée de résistance trop faible pour être utilisée.

[S]

[Serveur]
Système informatique destiné à fournir des services à des utilisateurs connectés.
Egalement, un ordinateur dédié à l'administration d'un service sur un réseau informatique. Il gère l'accès aux ressources et aux périphériques ainsi que les connexions des différents utilisateurs. Ainsi, un serveur de fichiers prépare la place mémoire pour des fichiers, un serveur d'impression gère et exécute les sorties sur imprimantes du réseau, enfin un serveur d'applications rend disponible sur son disque dur les programmes pouvant être appelés à travers le réseau.

[Serveur proxy]
Le serveur proxy est un serveur intermédiaire entre un internaute et Internet. Il joue le plus souvent un rôle de cache Web, et améliore de façon significative les performances d'Internet. De nombreuses organisations de grande taille ou fournisseurs d'accès Internet ont d'ores et déjà appliqué cette solution. Chaque page, image ou logo importé par un membre de l'organisation est stocké dans un cache (antémémoire) et sera immédiatement disponible pour un autre membre de l'organisation.
Outre un service de cache, un serveur proxy peut également jouer le rôle d'un outil de contrôle administratif (dans le cas d'une entreprise, qui veut surveiller les connexions de ses employés par exemple) ou de serveur firewall protégeant le réseau local des attaques extérieures.
Généralement, un serveur proxy est associé à ou fait partie du serveur passerelle qui sépare le réseau local de l'entreprise du réseau extérieur (Internet). Dans ce cas, il n'est plus nécessaire que chaque membre de l'organisation en amont du serveur proxy dispose de sa propre adresse IP, parce qu'ils n'ont plus un accès direct à Internet (rôle de NAT). Ainsi, un serveur proxy reçoit les requêtes des utilisateurs pour un certain service (tel qu'une demande d'une page Web). Si cette requête satisfait les conditions de filtrage, le serveur proxy, en supposant qu'il est aussi un serveur de cache, regardera dans sa mémoire cache locale contenant les pages précédemment téléchargées. S'il trouve la page voulue dans le cache, il l'a retourne directement à l'utilisateur, sans avoir besoin de transférer la requête vers Internet. Sinon, le serveur proxy, agissant comme un client de la part de l'utilisateur, utilise une de ses propres adresses IP pour émettre une requête demandant la page au serveur qui l'héberge sur Internet. Quand la page est retournée, le serveur proxy la relie à la requête originelle et la transmet à l'utilisateur qui en a fait la demande.

[Service de noms de domaine (Domain Name Service DNS)]
Le Service de noms de domaine (Domain Name Service DNS) est un mécanisme permettant d'attribuer (et, par la suite, de traduire) des noms, significatifs et faciles à se rappeler, aux ordinateurs identifiés par une adresse IP. Ces noms ont la forme de <nom>.<domaine>, où <nom> est une chaîne de caractères constituée d'une ou de plusieurs sous-chaînes séparées par des points.

[Serveur DNS (Domain Name Server)]
Serveur de nom de domaine. C'est lui qui convertit des noms de domaines en adresses IP. Le serveur DNS utilisé par un internaute se trouve généralement chez son fournisseur d'accès.

[Shareware]
Logiciel qui peut être importé d'Internet. Il est habituellement possible de l'importer pour un essai gratuit, mais pour l'utiliser légalement il faut payer un montant modeste à son développeur. Lorsqu'il est possible de l'importer et de l'employer tout à fait gratuitement, on parle de *freeware*.

[Signature numérique]
Données ajoutées à un ensemble de données, ou transformation cryptographique d'un ensemble de données, qui permet au récepteur de l'ensemble de données de prouver son origine et son intégrité et qui protège contre la fraude.
Une signature électronique est souvent une chaîne de caractères ajoutée à un message et qui garantit son intégrité en le cryptant (lui ou un condensé) avec la clé privée du signataire.
Quiconque reçoit le message signé peut vérifier s'il a été modifié en décryptant simplement la signature avec la clé publique de l'expéditeur et en comparant la chaîne de caractères décryptée avec le message original ou son condensé.

[Site Web]
Ensemble cohérent formé de fichiers de toute nature, qu'il s'agisse de documents HTML, de ressources audio-visuelles ou d'autres types de fichiers, identifié par une URL (celle de sa page d'accueil) et hébergé par un serveur Web. Chaque site est au moins composé d'une ou de plusieurs pages HTML reliées entre elles par des hyperliens.

[SMTP]
Simple Mail Transfer Protocol est un protocole applicatif reposant sur TCP/IP utilisé pour envoyer et recevoir des e-mails. Cependant, du fait de sa capacité limitée de créer une file d'attente de messages à la réception, il est habituellement utilisé avec un ou deux autres protocoles, POP3 ou Internet Message Access Protocol (IMAP), qui permettent à l'utilisateur de sauvegarder ses messages dans un serveur de messagerie contenant sa boîte aux lettres et de les télécharger périodiquement à partir de ce serveur. Typiquement, un utilisateur se sert d'un programme employant SMTP pour envoyer des e-mails et utilisant soit POP3 soit IMAP pour la réception des messages qui ont été reçu pour lui par un serveur local.

[Sniffing]
Le logiciel de sniffing permet de contrôler le trafic sur un réseau et d'y lire tous les paquets de données, en produisant donc un texte clair de toutes les communications non cryptées. La forme la plus simple de sniffing peut être utilisée sur un PC ordinaire connecté à un réseau, en employant un logiciel aisément disponible.

[Socks]
Socks est un protocole proxy conçu à l'origine pour permettre à des postes de travail d'un réseau privé local d'accèder à des réseaux publics comme Internet, tout en protégeant le réseau privé des accès externes. Typiquement les clients Socks passent au-travers d'un serveur Socks situé avant le coupe-feu pour se connecter à un serveur d'applications sur Internet. Le serveur Socks relaie les données entre le client et le serveur d'applications. Du point de vue du serveur d'application, le serveur Socks est le véritable client.
Il y a deux versions du protocole Socks, version 4 et version 5.
Le protocole Socksv4 a trois fonctions : il génère des requêtes de connection, établit des circuits proxy et relaie les données applicatives. Le protocole Socksv5 ajoute l'authentification et le support d'UDP.
Socksv5 est un protocole générique pour les applications communicantes TCP/IP, qui fait l'objet d'un standard IETF (Internet Engineering Task Force) : RFC 1928. Le protocole Socks fournit une architecture flexible pour développer des communications sûres en intégrant facilement d'autres technologies de sécurité.
Socks repose sur deux composants, le serveur Socks et le client Socks. Le serveur Socks est mis en oeuvre au niveau de la couche application, tandis que le client Socks est mis en oeuvre entre les couches application et transport. L'objectif principal du protocole est de permettre à des systèmes situés d'un côté d'un serveur Socks d'accéder à des systèmes situés de l'autre côté d'un serveur Socks, sans avoir besoin de connexion IP directe.

[Spamming (ou spam)]
Terme générique désignant un ensemble de pratiques abusives affectant des fonctionnalités de l'Internet. Dans son acceptation la plus courante, le spamming consiste en l'envoi massif et répété de courriers électroniques non sollicités, le plus souvent à caractère publicitaire.
C'est souvent le cas d'e-mails commerciaux envoyés en grand nombre sans prise en compte de leur pertinence auprès du public destinataire ou envoyés à des listes d'adresses e-mails créées par des compagnies spécialisées dans la création de listes de distribution. Pour les receveurs, ils sont souvent perçus comme des e-mails « poubelle » qui encombrent inutilement sa boîte aux lettres (et le réseau).
Le terme trouve son origine dans un sketch des Monty Pythons où les protagonistes répètent en permanence le terme saucisson - "spam" - au point de devenir absolument horripilants.

[SSH]
Secure Shell est un protocole de communication sécurisée permettant l'accès distant à des machines Unix (notamment pour les commandes telles que rlogin, rsh et rcp). SSH permet de pallier les faiblesses de sécurité des accès distants aux systèmes Unix (ex. : telnet, X11) en fournissant les services de sécurité essentiels : authentification du serveur, confidentialité des flux (notamment des mots de passe).
SSH repose sur la technique de cryptographie asymétrique RSA. SSH utilise les algorithmes symétriques IDEA (par défaut), Blowfish et DES pour la confidentialité des données.

[SSL]
Secure Socket Layer est un protocole de communication sécurisée fournissant des services de sécurité basés sur les techniques de cryptographie symétriques (DES, 3-DES, RCx) et asymétriques (RSA) :

- Authentification (unidirectionnelle ou mutuelle).
- Confidentialité des données.
- Intégrité des données.

SSL a été développé par Netscape (1ère version testée en interne, version 2.0 publiée en 1994, version actuelle 3.0 publié dans un draft IETF en 1996). Aujourd'hui, de nombreux éditeurs de logiciels le supportent. SSL est en cours de standardisation par l'IETF sous le nom TLS (Transport Layer Security).
SSL est une couche supplémentaire au système OSI située entre la couche transport (TCP) et la couche des services applicatifs.

SSL est composé de deux niveaux :

- 1er niveau : protocole au dessus de TCP/IP : Record Layer Protocol.
- 2nd niveau : 3 sous-protocoles : Handshake protocol, Change cipher protocol, Alert protocol.

Comme SSL est un protocole de communication indépendant du protocole de communication de plus haut niveau qui repose sur lui, il est donc possible de porter les logiciels de communications usuels (FTP, telnet, HTTP, etc.) sur SSL sans grande modification, et de façon quasiment transparente pour l'utilisateur. SSL peut alors négocier la méthode de chiffrement à utiliser, authentifier les acteurs de la communication et chiffrer au vol tout ce qui transite par son canal.
Il existe des patchs pour intégrer SSL aux logiciels de communications usuels et la majorité des navigateurs le supportent nativement.

[T]

TCP]
Transmission Control Protocol est un protocole utilisé conjointement avec le Internet Protocol (IP) pour envoyer des données, sous forme de messages, entre ordinateurs sur un réseau comme Internet. TCP permet à deux ordinateurs d'établir une liaison entre eux et de contrôler l'émission et la réception de messages quel qu'en soit le contenu (mais aussi de gérer la congestion du réseau le cas échéant), tandis que IP lui a pour but de régir l'attribution des adresses IP et la manière dont les informations circulent depuis l'émetteur jusqu'au destinataire. Cette série de règles régit le fonctionnement d'un réseau, indépendamment des différences entre les ordinateurs le composant.

[TCP/IP]
TCP/IP est le couple de protocoles standard à la base des communications sur Internet. Ils peuvent aussi être utilisés comme protocoles de communication dans un réseau privé hétérogène. Tout ordinateur connecté à Internet est équipé de plusieurs programmes implémentant ces deux protocoles.
TCP/IP se situe sur deux couches : TCP occupe la couche la plus haute, la couche transport (elle contrôle la division des messages en paquets de taille plus petite qui sont envoyés sur Internet et reçus par la couche TCP de la destination qui les réassemble pour reformer le message originel) et IP, la couche la plus basse, la couche réseau (qui gère l'adressage de chaque paquet afin qu'il arrive à la bonne destination).
Chaque routeur sur le réseau détermine la « meilleure » route à suivre pour que le paquet atteigne sa destination. Même si certains paquets issus du même message suivent des routes différentes des autres, ils seront réassemblés correctement à la destination.
Pour être en mesure d'échanger des paquets entre différents ordinateurs, TCP/IP exige de spécifier les trois valeurs suivantes : une adresse IP, un masque de sous-réseau et une passerelle (routeur) par défaut.

[Technique cryptographique symétrique]
Technique cryptographique qui utilise la même clé secrète pour la transformation cryptographique de l'émetteur ou du destinataire. Sans connaissance de la clé secrète, il est impossible de calculer la transformation cryptographique de l'émetteur ou du destinataire.

[Traduction d'adresse]
La traduction d'adresse (NAT : Network Address Translation) est un mécanisme initialement mis en oeuvre pour lutter contre la pénurie d'adresses IPv4.
Le principe consiste, pour un paquet IP donné, à modifier son adresse IP source ou destination. En particulier, certains équipements réseaux ou de sécurité (ex. : firewalls) réalisent de la traduction d'adresses pour masquer le plan d'adressage d'une entreprise.
La traduction d'adresse peut-être réalisée de deux manières :

- 1 pour 1 : une adresse est systématiquement masquée par une autre adresse unique.
- N pour 1 : N adresses IP sont masquées par la même adresse IP. Les paquets sont alors différencier par l'affectation de numéro de ports particuliers.

[Traitement de données à caractère personnel]
Désigne toute opération ou ensemble d'opérations, effectuées ou non à l'aide de procédés automatisés et appliquées à des données à caractère personnel. Les opérations concernées sont notamment : la collecte, l'enregistrement, l'organisation, la modification, l'extraction, la consultation, l'utilisation, la diffusion des données.

[Tunneling]
Connexion IP de deux ordinateurs isolés du reste du réseau par une méthode de chiffrement des paquets IP, par opposition à une connexion IP classique qui transmet ses paquets IP en clair.
Le principe de tunneling consiste à utiliser un réseau public non sécurisé (ex. : Internet) comme élément/extension d'un réseau privé sécurisé. On utilise ainsi les capacités de télécommunication de l'Internet en ajoutant une couche de sécurité afin d'établir un VPN. Les protocoles IPSEC, SSL ou SSH sont souvent utilisés pour établir cette couche supplémentaire de sécurité.

[U]

[UDP]
Le protocole UDP (User Datagram Protocol) est un protocole non orienté connexion de la couche transport du modèle TCP/IP. Ce protocole est très simple étant donné qu'il ne fournit pas de contrôle d'erreurs, contrairement à TCP.

[URL (Uniform Ressource Locator)]
Une URL est l'adresse d'une ressource (généralement un fichier) accessible sur Internet.
Le type de la ressource dépend du protocole applicatif utilisé pour l'obtenir. Utilisant le HTTP, la ressource peut être une page HTML une image, un programme, une Applet Java ou n'importe quel autre fichier supporté par ce protocole.
L'URL s'écrit de la façon suivante :

protocol://server/directory/document.

Elle peut être complétée par des paramètres précédés d'un "?".
L'URL contient donc le nom du protocole requis pour accéder à la ressource désignée, un nom de domaine qui identifie l'ordinateur spécifique qui héberge la ressource et une description hiérarchique de la localisation du fichier sur l'ordinateur ainsi que le nom de ce fichier.

[V]

[Variables d'environnement]
Données automatiquement transmises au site visité par le logiciel de navigation utilisé par l'internaute, à savoir : l'adresse TCP/IP, la marque et la version du navigateur ainsi que du système d'exploitation, la langue utilisée par l'internaute, la dernière page Web consultée.

[VB]
Visual Basic est un environnement de programmation de Microsoft dans lequel un développeur utilise une interface graphique pour choisir et modifier des sections de code présélectionnées écrites dans langage de programmation propriétaire, accessible aux débutants, tout usage. Parce que le Visual Basic est facile à apprendre et que l'on peut rapidement créer du code, il est parfois utilisé pour prototyper une application qui sera écrite plus tard dans un langage plus difficile mais plus efficace. Visual Basic est aussi largement utilisé pour écrire des programmes finaux. Plusieurs millions de programmeurs utilisent ce langage.

VBScript]

Le VBScript est un langage de script interprété mis au point par Microsoft. C'est un sous-ensemble du langage de programmation Visual Basic. VBScript peut être comparé à d'autres langages de scripts conçus pour le Web, dont le JavaScript de Netscape. En général, les langages de scripts sont plus faciles et plus rapides à coder que les langages compilés et plus structurés comme le C ou le C++. C'est donc l'idéal pour les petits programmes aux capacités limitées, qui peuvent être réutilisés ou liés avec des programmes compilés préexistants. VBScript est à coup sûr la réponse de Microsoft au populaire JavaScript.

[VPN]

Un Virtual Private Network (Réseau Privé Virtuel) est un réseau de données qui utilise les moyens de télécommunications d'un réseau public en ajoutant des services de sécurité et des protocoles de tunneling. Internet est en général utilisé pour établir un VPN pour des raisons de coûts.

[W]

[W3C]

Le World Wide Web Consortium (W3C) est une organisation à but non lucratif créée en octobre 1994 « pour conduire le World Wide Web vers son plein potentiel en développant des protocoles communs qui promouvoient son évolution et garantissent son interopérabilité ». W3C a plus de 400 organisations membres réparties dans le monde entier.

[Web bug (ou pixel espion)]

Image incorporée abusivement dans une page Web ou dans un message de courrier électronique et conçue pour identifier celui qui lit la page Web ou le courrier électronique en question. Le pixel espion est invisible et pratiquement indétectable (sauf en accédant au code source de la page), car sa taille n'est que d'un pixel sur un pixel. Ils sont très souvent utilisés à des fins de pistage, de profilage ou de mesure d'audience par certains sites tiers (appartenant généralement à des sociétés de cybermarketing) et/ou pour recueillir des informations sur les internautes. Pour ce faire, un Web bug est généralement lié à un cookie ou à un script écrit en JavaScript.

[Webmail]

Systèmes de e-mail utilisant les pages Web comme interface (ex: Yahoo!, HotMail, etc.). Le Webmail est accessible de n'importe où, et l'utilisateur ne doit pas établir de connexion avec un ISP particulier, comme dans un autre compte e-mail ordinaire.

[Webmaster]

Responsable d'un site Web, parfois désigne également son créateur.

[World Wide Web (ou Web)]

Le Web est un système d'applications client/serveur qui transfère une grande variété de documents, dont les pages HTML constituent un des meilleurs exemples. Les transferts d'informations utilisent généralement le protocole HTTP. L'accès au Web est rendu possible à un utilisateur grâce au navigateur. le Web est le domaine de l'Internet le plus populaire en raison de son caractère multimédia (texte, son et image), de ses moteurs de recherche et du fait qu'il permet d'utiliser l'hypertexte.

[X]

[XML]
Méta-langage extensible dérivé de SGML (Standard Generalized Markup Language) permettant de structurer des données.
eXtensible Markup Language est une manière flexible de créer des formats de documents et de partager à le fois le format et les données sur le Web ou dans n'importe quel réseau. XML peut être utilisé par une personne, un groupe de personnes ou une organisation pour partager de l'information d'une manière consistante. XML est une recommandation du W3C.

LISTE DES ACRONYMES

ADSL	Asynchronous Digital Subscriber Line
API	Application Programming Interface
CERT	Computer Emergency Response Team
CGI	Commun Gateway Interface
DHCP	Dynamic Host Configuration Protocol
DNS	Domain Name Service
DNSSEC	Domain Name Service Security Extensions
DoS	Denial of Service
FAQ	Frequently Asked Questions
FTP	File Transfer Protocol
HTML	Hypertext Markup Language
HTTP	Hypertext Transport Protocol
HTTPS	Secure Hypertext Transport Protocol
IETF	Internet Engineering Task Force
IMAP	Internet Message Access Protocol
IP	Internet Protocol
IPSEC	Internet Protocol SECurity
ISDN	Integrated Services Digital Network
ISOC	Internet Society
ISP	Internet Service Provider
IRC	Internet Relay Chat
LAN	Local Area Network
MIME	Multi-Purpose Internet Mail Extensions
NAT	Network Address Translation
NNTP	Network News Transfer Protocol
OSI	Open Systems Interconnection
POP	Post Office Protocol
PPP	Point-to-Point Protocol
RFC	Request for Comments
SMTP	Simple Mail Transfer Protocol
SSH	Secure SHell
SSL	Secure Socket Layer
TCP	Transmission Control Protocol
UDP	User Datagramme Protocol
URL	Uniform Resource Locator
VPN	Virtual Private Network
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	eXtensible Markup Language

MS	Microsoft
MSIE	Microsoft Internet Explorer

ANNEXE B : Implémentation d'un filtre de protection de la vie privée pour Jigsaw

classes de Jigsaw modifiées :

- HttpRequestMessage (ligne 450 et 454 - remplacement de HttpAcceptCharset par HttpAcceptLanguage)

- ProxyProp (insertion du filtre PrivacyProtectionFilter)

- HttpAccept (constructeur public)

remplacement de

```
HttpAccept(boolean isValid, MimeType type, double quality) {
        this.isValid = isValid;
        setMimeType(type);
        this.quality = quality;
}
```

par

```
public HttpAccept(boolean isValid, MimeType type, double quality) {
    this.isValid = isValid;
    setMimeType(type);
    this.quality = quality;
}
```

- HttpCookieList (constructeur public + ajout méthode getCookiesLength() + ajout méthode addCookie(HttpCookie cookie))

remplacement de

```
HttpCookieList(HttpCookie c[]) {
        this.isValid = true;
        this.cookies = new Vector(8);
        if ( c != null ) {
                for (int i = 0 ; i < c.length ; i++)
                cookies.addElement(c[i]);
        }
 }
```

par

```
public HttpCookieList(HttpCookie c[]) {
        this.isValid = true;
        this.cookies = new Vector(8);
        if ( c != null ) {
            for (int i = 0 ; i < c.length ; i++)
                cookies.addElement(c[i]);
        }
}
```

et ajout de

```
public int getCookiesLength() {
        return cookies.size();
}
```

et de

```
public void addCookie(HttpCookie cookie) {
        validate();
        cookies.addElement(cookie);
}
```

- HttpSetCookieList (constructeur public + ajout méthode getSetCookiesLength())

remplacement de

```
HttpSetCookieList() {
        this.isValid    = false;
        this.setcookies = new Vector(2);
}
```

par

```
public HttpSetCookieList() {
        this.isValid    = false;
        this.setcookies = new Vector(2);
}
```

et ajout de

```
public int getSetCookiesLength() {
        return setcookies.size();
}
```

```
// PrivacyProtectionFilter.java
// $Id: PrivacyProtectionFilter.java,v 0.4 2002/08/30 St. Liénart Exp $
// (c) COPYRIGHT FUNDP, 2002.


package org.w3c.www.protocol.http.privacy;

import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.io.InputStreamReader;
import java.io.OutputStreamWriter;
import java.io.PrintStream;
import java.io.Writer;

import java.util.Date;
import java.util.Enumeration;
import java.util.Vector;

import java.io.File;
import java.net.URL;

import org.w3c.util.ObservableProperties;
import org.w3c.util.PropertyMonitoring;

import org.w3c.www.protocol.http.HttpException;
import org.w3c.www.protocol.http.HttpManager;
import org.w3c.www.protocol.http.PropRequestFilter;
import org.w3c.www.protocol.http.PropRequestFilterException;
import org.w3c.www.protocol.http.Reply;
import org.w3c.www.protocol.http.Request;
import org.w3c.www.mime.MimeType;

import org.w3c.www.http.HTTP;
import org.w3c.www.http.HttpEntityMessage;
import org.w3c.www.http.HttpCookie;
import org.w3c.www.http.HttpCookieList;
import org.w3c.www.http.HttpFactory;
import org.w3c.www.http.HttpMessage;
import org.w3c.www.http.HttpReplyMessage;
import org.w3c.www.http.HttpRequestMessage;
import org.w3c.www.http.HttpSetCookie;
import org.w3c.www.http.HttpSetCookieList;
import org.w3c.www.http.HttpAccept;
import org.w3c.www.http.HttpAcceptLanguage;


/**
 * Client side PrivacyProtectionFilter :
 * @author Stéphane Liénart <lienart@info.fundp.ac.be>
 */

public class PrivacyProtectionFilter implements PropRequestFilter,
                                                PropertyMonitoring {

    /**
     * Properties - Using the black list option ?
     */
    public static final String PRIVACY_BLACKLIST_P              =
            "org.w3c.protocol.http.privacy.blacklist";
            //"Use.BlackList.Option";

    /**
     * Properties - the file containing the black list
     */
    public static final String PRIVACY_BLACKLIST_FILE_P         =
            "org.w3c.protocol.http.privacy.blacklist.file";
            //"Location.of.the.BlackList.File";
```

```
77      /**
78       * Properties - Jigsaw proxy manage himself cookies for his clients ?
79       */
80      public static final String PRIVACY_COOKIES_P                =
81              "org.w3c.protocol.http.privacy.cookies";
82              //"Use.Jigsaw's.Cookies.Manager";
83
84      /**
85       * Properties - the file containing the Jigsaw's cookies
86       */
87      public static final String PRIVACY_COOKIES_FILE_P           =
88              "org.w3c.protocol.http.privacy.cookies.file";
89              //"Location.of.the.Cookies.File";
90
91      /**
92       * Properties - Using black & white lists of cookie domains ?
93       */
94      public static final String PRIVACY_COOKIELISTS_P            =
95              "org.w3c.protocol.http.privacy.cookielists";
96              //"Use.Black.and.White.Lists.of.Cookie";
97
98      /**
99       * Properties - the file containing the black & white lists of
100      * cookie domains
101      */
102     public static final String PRIVACY_COOKIELISTS_FILE_P       =
103             "org.w3c.protocol.http.privacy.cookielists.file";
104             //"Location.of.the.Black.and.White.Lists.of.Cookie.File";
105
106     /**
107      * Properties - Keep the VIA header ?
108      */
109     public static final String PRIVACY_VIA_P                    =
110             "org.w3c.protocol.http.privacy.via";
111             //"Transmit.Via.Header";
112
113     /**
114      * Properties - Modify the cookie maxage field of cookies ?
115      */
116     public static final String PRIVACY_COOKIE_MAXAGE_SET_P      =
117             "org.w3c.protocol.http.privacy.cookie_maxage_set";
118             //"Modify.Cookie.Maxage";
119
120     /**
121      * Properties - the new value of cookie maxage field (in days)
122      */
123     public static final String PRIVACY_COOKIE_MAXAGE_VALUE_P    =
124             "org.w3c.protocol.http.privacy.cookie_maxage_value";
125             //"Set.the.new.Cookie.Maxage.Value";
126
127     /**
128      * Properties - Keep the ACCEPT header ?
129      */
130     public static final String PRIVACY_ACCEPT_P                 =
131             "org.w3c.protocol.http.privacy.accept";
132             //"Transmit.Accept.Header";
133
134     /**
135      * Properties - Keep the ACCEPT_LANGUAGE header ?
136      */
137     public static final String PRIVACY_ACCEPT_LANGUAGE_P        =
138             "org.w3c.protocol.http.privacy.accept_language";
139             //"Transmit.Accept.Language.Header";
140
141     /**
142      * Properties - Accept cookies from Clients to Servers ?
143      */
144     public static final String PRIVACY_ACCEPT_COOKIE_CS_P       =
145             "org.w3c.protocol.http.privacy.accept_cookie_cs";
146             //"Accept.Cookie.Stream.from.Clients.to.Servers";
147
148
149
150
151
152
```

```
    /**
     * Properties - Accept cookies from Servers to Clients ?
     */
    public static final String PRIVACY_ACCEPT_COOKIE_SC_P       =
            "org.w3c.protocol.http.privacy.accept_cookie_sc";
            //"Accept.Cookie.Stream.from.Servers.to.Clients";

    /**
     * Properties - Accept session cookies ?
     */
    public static final String PRIVACY_ACCEPT_SESSION_COOKIE_P  =
            "org.w3c.protocol.http.privacy.accept_session_cookie";
            //"Accept.Session.Cookies";

    /**
     * Properties - Keep the FROM header ?
     */
    public static final String PRIVACY_FROM_P                   =
            "org.w3c.protocol.http.privacy.from";
            //"Transmit.From.Header";

    /**
     * Properties - Keep the REFERER header ?
     */
    public static final String PRIVACY_REFERER_P                =
            "org.w3c.protocol.http.privacy.referer";
            //"Transmit.Referer.Header";

    /**
     * Properties - Modify the USER_AGENT header
     */
    public static final String PRIVACY_USER_AGENT_P             =
            "org.w3c.protocol.http.privacy.user_agent";
            //"Modify.User_Agent.Header";

    private static final String defaultFileName_h   = "blacklist";
    private static final String defaultFileName_c   = "cookies";
    private static final String defaultFileName_cl  = "white_and_black_cookie_lists";

    protected boolean blacklist             = true;
    protected boolean cookies               = true;
    protected boolean cookielists           = true;
    protected boolean via                   = false;
    protected boolean accept                = false;
    protected boolean accept_language       = false;
    protected boolean accept_cookie_cs      = false;
    protected boolean accept_cookie_sc      = false;
    protected boolean accept_session_cookie = true;
    protected boolean maxage_set            = false;
    protected int maxage_value              = -1;  //-1 for session cookie
    protected boolean from                  = false;
    protected boolean referer               = false;

    /**
     * the new USER_AGENT header (<strong>null</strong> to
     * suppress this header)
     */
    protected String user_agent             = "Mozilla/3.0";

    private static File blacklistfile       = null;
    private static File cookiesfile         = null;
    private static File cookielistsfile     = null;

    protected HttpManager manager       = null;
    protected ObservableProperties props = null;
    private static BlackList root_hl    = null;
    private static CookieLists root_cl  = null;
    private static DomainTree root_dt   = null;

    static {
        root_dt = new DomainTree();
        root_hl = new BlackList();
        root_cl = new CookieLists();
    }
```

```
229     /**
230      * The request pre-processing hook.
231      * Before each request is launched, all filters will be called back
232      * through
233      * this method. They will generally set up additional request header
234      * fields to enhance the request.
235      * @param request The request that is about to be launched.
236      * @return An instance of Reply if the filter could handle the request,
237      * or <strong>null</strong> if processing should continue normally.
238      * @exception HttpException If the filter is supposed to fulfill the
239      * request, but some error happened during that processing.
240      */
241
242     public Reply ingoingFilter(Request request) throws HttpException  {
243
244         if (blacklist && (! root_hl.isEmpty()))
245         {
246             if (root_hl.matchBlackList(request.getURL()))
247             {// request's URL in blacklist file
248                 Reply reply = request.makeReply(HTTP.ACCEPTED);
249                 reply.setContent("<h1>Access forbidden</h1>\n"
250                     + "<p>Access to "+request.getURL()
251                     + " is forbidden by the Jigsaw's black list.");
252
253                 return reply;
254             }
255         }
256
257         if (! via)
258         {
259             request.setVia(null);
260         }
261
262         if (! from)
263         {
264             request.setFrom(null);
265         }
266
267         if (! referer)
268         {
269             request.setReferer(null);
270         }
271
272         if (! accept)
273         {
274             MimeType mt = new MimeType("*","*");
275             HttpAccept ha = new HttpAccept(true, mt, 1.0);
276             HttpAccept [] hal = new HttpAccept[1];
277             hal[0] = ha;
278             request.setAccept(hal);
279         }
280
281         if (! accept_language)
282         {
283             request.setAcceptLanguage(null);
284         }
285
286         request.setUserAgent(user_agent);
287
288         if (cookies)
289         {
290             HttpCookieList cooklist = root_dt.getCookies(request.getURL());
291             request.setCookie(cooklist);
292             request = this.manageCookie(request);
293
294             return null;
295         } else  {
296             request = this.manageCookie(request);
297             return null;
298         }
299
300     }
301
302
303
304
```

```
    protected Request manageCookie(Request request) {

        String name = "";
        String domaine = "";
        HttpCookieList cooklist;
        HttpCookieList cooklistbis = new HttpCookieList(null);
        HttpCookie [] cooks;

        cooklist = request.getCookie();
        if ((cooklist != null) && (cooklist.getCookiesLength() != 0))
        {
            if (cookielists && (! root_cl.isEmpty("+")))
            {
                cooks = cooklist.getCookies();
                for (int i = 0 ; i < cooks.length ; i++) {
                    domaine =  cooks[i].getDomain();
                    if (root_cl.matchCookieLists(domaine, "+"))
                    {
                        name = cooks[i].getName();
                        cooklistbis.addCookie(cooks[i]);
                        cooklist.removeCookie(name);
                    }
                }
                if (cooklist.getCookiesLength() != 0)
                {
                    request.setCookie(cooklist);
                } else {
                    if (cooklistbis.getCookiesLength() != 0)
                    {
                        request.setCookie(cooklistbis);
                        return request;
                    }
                }
            }

            if (cookielists && (! root_cl.isEmpty("-")))
            {
                cooklist = request.getCookie();
                if ((cooklist != null) && (cooklist.getCookiesLength() != 0))
                {
                    cooks = cooklist.getCookies();
                    for (int i = 0 ; i < cooks.length ; i++) {
                        domaine =  cooks[i].getDomain();
                        if (root_cl.matchCookieLists(domaine, "-"))
                        {
                            name = cooks[i].getName();
                            cooklist.removeCookie(name);
                        }
                    }
                    if (cooklist.getCookiesLength() != 0)
                    {
                        request.setCookie(cooklist);
                    } else {
                            request.setCookie(null);
                            return request;
                    }
                }
            }

            cooklist = request.getCookie();
        if ((cooklist != null) && (cooklist.getCookiesLength() != 0))
            {
                cooks = cooklist.getCookies();

                if (accept_cookie_cs)
                {
                    if (! accept_session_cookie)
                    {
                        for (int i = 0 ; i < cooks.length ; i++) {
                            name = cooks[i].getName();
                            if (name.endsWith("_SESSION"))
                            {
                                cooklist.removeCookie(name);
                            }
                        }
                        if (cooklist.getCookiesLength() != 0)
```

```
                        {
                            request.setCookie(cooklist);
                        } else {
                            //cooklist.addCookie("bidon", "bidon"); //WAFER
                            //request.setCookie(cooklist);
                            request.setCookie(null);
                            return request;
                        }
                    }
                }
                if (! accept_cookie_cs)
                {
                    if (! accept_session_cookie)
                    {
                        //cooklist.addCookie("bidon", "bidon"); //WAFER
                        //request.setCookie(cooklist);
                        request.setCookie(null);
                        return request;
                    }
                    if (accept_session_cookie)
                    {
                        for (int i = 0 ; i < cooks.length ; i++) {
                            name =  cooks[i].getName();
                            if (! name.endsWith("_SESSION"))
                            {
                                cooklist.removeCookie(name);
                            }
                        }
                        if (cooklist.getCookiesLength() != 0)
                        {
                            request.setCookie(cooklist);
                        } else {
                            //cooklist.addCookie("bidon", "bidon"); //WAFER
                            //request.setCookie(cooklist);
                            request.setCookie(null);
                            return request;
                        }
                    }
                }
            }

            if (cooklistbis.getCookiesLength() != 0)
            {
                cooklist = request.getCookie();
                if ((cooklist != null) && (cooklist.getCookiesLength() != 0))
                {
                    cooks = cooklistbis.getCookies();
                    for (int i = 0 ; i < cooks.length ; i++) {
                        cooklist.addCookie(cooks[i]);
                    }
                    request.setCookie(cooklist);
                } else { request.setCookie(cooklistbis); }
            }

            cooklist = request.getCookie();
        if ((cooklist != null) && (cooklist.getCookiesLength() != 0))
            {
                cooks = cooklist.getCookies();
                for (int i = 0 ; i < cooks.length ; i++) {
                    name =  cooks[i].getName();
                    if (name.endsWith("_SESSION"))
                    {
                        int j = name.lastIndexOf("_SESSION");
                        name = name.substring(j);
                        cooks[i].setName(name);
                    }
                }
                HttpCookieList cooklistter = new HttpCookieList(cooks);
                request.setCookie(cooklistter);
            }

            return request;
        } else {return request; }

    }
```

```
    /**
     * The request post-processing hook.
     * After each request has been replied to by the target server (be it a
     * proxy or the actual origin server), each filter's outgoingFilter
     * method is called.
     * <p>It gets the original request, and the actual reply as a parameter,
     * and should return whatever reply it wants the caller to get.
     * @param request The original (handled) request.
     * @param reply The reply, as emited by the target server, or constructed
     * by some other filter.
     * @exception HttpException If the reply emitted by the server is not
     * a valid HTTP reply.
     */

    public Reply outgoingFilter(Request request, Reply reply)
    throws HttpException
    {
        if (cookies)
        {
            reply = this.manageSetCookie(reply);

            HttpSetCookieList list = reply.getSetCookie();
            if ((list != null) && (list.getSetCookiesLength() != 0))
            {
                HttpSetCookie [] cooks = list.getSetCookies();
                if (cooks != null) {
                    int i = 0;
                    while (i < cooks.length) {
                        if (cooks[i].getDomain() == null)
                        { cooks[i].setDomain(request.getURL().getHost()); }
                        root_dt.insertCookie(cooks[i++]);
                    }
                }
            }

            reply.setSetCookie(null);
            return reply;

        } else {
            reply = this.manageSetCookie(reply);
            return reply;
        }
    }

    protected Reply manageSetCookie(Reply reply) {

        String name = "";
        String domaine = "";
        HttpSetCookieList cooklist;
        HttpSetCookieList cooklistbis = new HttpSetCookieList(null);
        HttpSetCookie [] cooks;

        cooklist = reply.getSetCookie();
        if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
        {
            cooks = cooklist.getSetCookies();
            for (int i = 0 ; i < cooks.length ; i++) {
                if (cooks[i].getMaxAge() == -1)
                {
                    name = cooks[i].getName();
                    name = name.concat("_SESSION");
                    cooks[i].setName(name);
                }
            }
            HttpSetCookieList cooklistter = new HttpSetCookieList(cooks);
            reply.setSetCookie(cooklistter);

            if (cookielists && (! root_cl.isEmpty("+")))
            {
                cooklist = reply.getSetCookie();
                if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
                {
                    cooks = cooklist.getSetCookies();
                    for (int i = 0 ; i < cooks.length ; i++) {
                        domaine =  cooks[i].getDomain();
                        if (root_cl.matchCookieLists(domaine, "+"))
```

```
533                         {
534                             name = cooks[i].getName();
535                             cooklistbis.addSetCookie(cooks[i]);
536                             cooklist.removeSetCookie(name);
537                         }
538                     }
539                     if (cooklist.getSetCookiesLength() != 0)
540                     {
541                         reply.setSetCookie(cooklist);
542                     } else {
543                         if (cooklistbis.getSetCookiesLength() != 0)
544                         {
545                             reply.setSetCookie(cooklistbis);
546                             return reply;
547                         }
548                     }
549                 }
550             }
551
552             if (cookielists && (! root_cl.isEmpty("-")))
553             {
554                 cooklist = reply.getSetCookie();
555                 if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
556                 {
557                     cooks = cooklist.getSetCookies();
558                     for (int i = 0 ; i < cooks.length ; i++) {
559                         domaine =  cooks[i].getDomain();
560                         if (root_cl.matchCookieLists(domaine, "-"))
561                         {
562                             name = cooks[i].getName();
563                             cooklist.removeSetCookie(name);
564                         }
565                     }
566                     if (cooklist.getSetCookiesLength() != 0)
567                     {
568                         reply.setSetCookie(cooklist);
569                     } else {
570                             reply.setSetCookie(null);
571                             return reply;
572                     }
573                 }
574             }
575
576             if (accept_cookie_sc)
577             {
578                 if (! accept_session_cookie)
579                 {
580                     cooklist = reply.getSetCookie();
581                 if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
582                     {
583                         cooks = cooklist.getSetCookies();
584                         for (int i = 0 ; i < cooks.length ; i++) {
585                             if (cooks[i].getMaxAge() == -1)
586                             {
587                                 name = cooks[i].getName();
588                                 cooklist.removeSetCookie(name);
589                             }
590                         }
591                         if (cooklist.getSetCookiesLength() != 0)
592                         {
593                             reply.setSetCookie(cooklist);
594                         } else {
595                             reply.setSetCookie(null);
596                             return reply;
597                         }
598                     }
599                 }
600
601                 if (maxage_set)
602                 {
603                     cooklist = reply.getSetCookie();
604                     if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
605                     {
606                         cooks = cooklist.getSetCookies();
607                         if (maxage_value >= 0)
608                         {
```

```
                            int cookieMaxage = 24*3600*maxage_value;
                            for (int i = 0 ; i < cooks.length ; i++) {
                                if (cooks[i].getMaxAge() != -1)
                                {
                                    cooks[i].setMaxAge(cookieMaxage);
                                }
                            }
                        }
                        if (maxage_value == -1)
                        {
                            for (int i = 0 ; i < cooks.length ; i++) {
                                if (cooks[i].getMaxAge() != -1)
                                {
                                    cooks[i].setMaxAge(-1);
                                }
                            }
                        }
                    HttpSetCookieList cooklistquat = new HttpSetCookieList(cooks);
                            reply.setSetCookie(cooklistquat);
                        }
                    }
                }

                if (! accept_cookie_sc)
                {
                    if (! accept_session_cookie)
                    {
                        reply.setSetCookie(null);
                        return reply;
                    }

                    if (accept_session_cookie)
                    {
                        cooklist = reply.getSetCookie();
                        if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
                        {
                            cooks = cooklist.getSetCookies();
                            for (int i = 0 ; i < cooks.length ; i++) {
                                if (cooks[i].getMaxAge() != -1)
                                {
                                    name = cooks[i].getName();
                                    cooklist.removeSetCookie(name);
                                }
                            }
                            if (cooklist.getSetCookiesLength() != 0)
                            {
                                reply.setSetCookie(cooklist);
                            } else {
                                reply.setSetCookie(null);
                                return reply;
                            }
                        }
                    }
                }

                if (cooklistbis.getSetCookiesLength() != 0)
                {
                    cooklist = reply.getSetCookie();
                if ((cooklist != null) && (cooklist.getSetCookiesLength() != 0))
                    {
                        cooks = cooklistbis.getSetCookies();
                        for (int i = 0 ; i < cooks.length ; i++) {
                            cooklist.addSetCookie(cooks[i]);
                        }
                        reply.setSetCookie(cooklist);
                    } else {
                        reply.setSetCookie(cooklistbis); }
                }

                return reply;
            } else { return reply; }
        }

        /**
         * An exception occured while talking to target server.
         * This method is triggered by the HttpManager, when the target server
```

```
685       * (which can be a proxy for that request) was not reachable, or some
686       * network error occured while emitting the request or reading the reply
687       * headers.
688       * @param request The request whose processing triggered the exception.
689       * @param ex The exception that was triggered.
690       * @return A boolean, <strong>true</strong> if that filter did influence
691       * the target server used to fulfill the request, and it has fixed the
692       * problem in such a way that the request should be retried.
693       * This filter doesn't handle exceptions.
694       * => return Always <strong>false</strong>.
695       */
696
697      public boolean exceptionFilter(Request request, HttpException ex) {
698          return false;
699      }
700
701      /**
702       * Synchronized any pending state into stable storage.
703       * If the filter maintains some in-memory cached state, this method
704       * should ensure that cached data are saved to stable storage.
705       */
706
707      public void sync() {
708
709          if (cookies)
710          {
711              root_dt.sync(cookiesfile);
712          }
713      }
714
715     /**
716       * PropertyMonitoring implementation - Commit property changes.
717       * @param name The name of the property that has changed.
718       * @return A boolean <strong>true</strong> if change was commited,
719       * <strong>false</strong> otherwise.
720       */
721
722      public boolean propertyChanged(String name) {
723
724          if ( name.equals(PRIVACY_BLACKLIST_FILE_P) ) {
725              blacklistfile = props.getFile(name, null);
726
727              if (blacklistfile == null) {
728                  blacklistfile = new File(defaultFileName_h);
729              }
730          if (blacklistfile.getAbsolutePath() == root_hl.blacklistfilename)
731              {
732                  return true;
733              }
734              root_hl.blacklistfilename = blacklistfile.getAbsolutePath();
735
736              root_hl.majB.destroy();
737
738              if (blacklistfile.exists()) {
739                  try {
740                          root_hl.loadBlackList();
741                          root_hl.launchThread();
742                      } catch (Exception ex) {
743                          ex.printStackTrace();
744                          System.out.println(ex.getMessage());
745                      }
746              }
747              return true;
748
749          } else if ( name.equals(PRIVACY_COOKIELISTS_FILE_P) ) {
750              cookielistsfile = props.getFile(name, null);
751
752              if (cookielistsfile == null) {
753                  cookielistsfile = new File(defaultFileName_cl);
754              }
755          if (cookielistsfile.getAbsolutePath() == root_cl.cookielistsfilename)
756              {
757                  return true;
758              }
759              root_cl.cookielistsfilename = cookielistsfile.getAbsolutePath();
```

```
            root_cl.majC.destroy();

            if (cookielistsfile.exists()) {
                try {
                        root_cl.loadCookieLists();
                        root_cl.launchThread();
                    } catch (Exception ex) {
                        ex.printStackTrace();
                        System.out.println(ex.getMessage());
                    }
            }
            return true;

        } else if ( name.equals(PRIVACY_BLACKLIST_P) ) {
            blacklist = props.getBoolean(name, blacklist);
            return true;
        } else if ( name.equals(PRIVACY_COOKIES_P) ) {
            cookies = props.getBoolean(name, cookies);
            return true;
        } else if ( name.equals(PRIVACY_COOKIELISTS_P) ) {
            cookielists = props.getBoolean(name, cookielists);
            return true;
        } else if ( name.equals(PRIVACY_VIA_P) ) {
            via = props.getBoolean(name, via);
            return true;
        } else if ( name.equals(PRIVACY_ACCEPT_P) ) {
            accept = props.getBoolean(name, accept);
            return true;
        } else if ( name.equals(PRIVACY_ACCEPT_LANGUAGE_P) ) {
            accept_language = props.getBoolean(name, accept_language);
            return true;
        } else if ( name.equals(PRIVACY_ACCEPT_COOKIE_CS_P)) {
            accept_cookie_cs = props.getBoolean(name, accept_cookie_cs);
            return true;
        } else if ( name.equals(PRIVACY_ACCEPT_COOKIE_SC_P) ) {
            accept_cookie_sc = props.getBoolean(name, accept_cookie_sc);
            return true;
        } else if ( name.equals(PRIVACY_ACCEPT_SESSION_COOKIE_P)) {
            accept_session_cookie = props.getBoolean(name, accept_session_cookie);
            return true;
        } else if ( name.equals(PRIVACY_COOKIE_MAXAGE_SET_P) ) {
            maxage_set = props.getBoolean(name, maxage_set);
            return true;
        } else if ( name.equals(PRIVACY_COOKIE_MAXAGE_VALUE_P)) {
            maxage_value = props.getInteger(name, maxage_value);
            return true;
        } else if ( name.equals(PRIVACY_FROM_P) ) {
            from = props.getBoolean(name, from);
            return true;
        } else if ( name.equals(PRIVACY_REFERER_P)) {
            referer = props.getBoolean(name, referer);
            return true;
        } else if ( name.equals(PRIVACY_USER_AGENT_P)) {
            user_agent = props.getString(name, user_agent);
            return true;
        } else {
        // nothing changed, everything is ok!
        return true; }
    }

    /**
     * Initialize this filter, using the provided manager.
     * During initialization, it is up to the filter to install itself
     * in the manager, by invoking the appropriate <code>setFilter</code>
     * method.
     * @param manager The HttpManager initializing the filter.
     * @exception PropRequestFilterException If the filter couldn't be
     * initialized properly.
     */

    public void initialize(HttpManager manager)
    throws PropRequestFilterException
    {
        this.manager = manager;
        props = manager.getProperties();
        props.registerObserver(this);
```

```
837
838         blacklist               = props.getBoolean(PRIVACY_BLACKLIST_P, true);
839         cookies                 = props.getBoolean(PRIVACY_COOKIES_P, true);
840         cookielists             = props.getBoolean(PRIVACY_COOKIELISTS_P, true);
841         via                     = props.getBoolean(PRIVACY_VIA_P, false);
842         accept                  = props.getBoolean(PRIVACY_ACCEPT_P, false);
843         accept_language         = props.getBoolean(PRIVACY_ACCEPT_LANGUAGE_P, false);
844         accept_cookie_cs        = props.getBoolean(PRIVACY_ACCEPT_COOKIE_CS_P, false);
845         accept_cookie_sc        = props.getBoolean(PRIVACY_ACCEPT_COOKIE_SC_P, false);
846         accept_session_cookie   = props.getBoolean(PRIVACY_ACCEPT_SESSION_COOKIE_P, true)
847         maxage_set              = props.getBoolean(PRIVACY_COOKIE_MAXAGE_SET_P, false);
848         maxage_value            = props.getInteger(PRIVACY_COOKIE_MAXAGE_VALUE_P, -1);
849         from                    = props.getBoolean(PRIVACY_FROM_P, false);
850         referer                 = props.getBoolean(PRIVACY_REFERER_P, false);
851         user_agent              = props.getString(PRIVACY_USER_AGENT_P, "Mozilla/3.0");
852         blacklistfile           = props.getFile(PRIVACY_BLACKLIST_FILE_P, null);
853         cookiesfile             = props.getFile(PRIVACY_COOKIES_FILE_P, null);
854         cookielistsfile         = props.getFile(PRIVACY_COOKIELISTS_FILE_P, null);
855
856         if (blacklistfile == null) {
857             blacklistfile = new File(defaultFileName_h);
858         }
859         root_hl.blacklistfilename = blacklistfile.getAbsolutePath();
860
861         if (blacklistfile.exists()) {
862             try {
863                     root_hl.loadBlackList();
864                     root_hl.launchThread();
865                 } catch (Exception ex) {
866                     System.out.println(ex.getMessage());
867                     throw new PropRequestFilterException(ex.getMessage());
868                 }
869         }
870
871         if (cookiesfile == null) {
872             cookiesfile = new File(defaultFileName_c);
873         }
874
875         if (cookiesfile.exists()) {
876             // load all the Cookie filter and register them in the manager
877             try {
878                 root_dt.loadCookies(cookiesfile);
879             } catch (FileNotFoundException ex) {
880                 System.out.println(ex.getMessage());
881                 throw new PropRequestFilterException(ex.getMessage());
882             }
883         }
884
885         if (cookielistsfile == null) {
886             cookielistsfile = new File(defaultFileName_cl);
887         }
888         root_cl.cookielistsfilename = cookielistsfile.getAbsolutePath();
889
890         if (cookielistsfile.exists()) {
891             try {
892                     root_cl.loadCookieLists();
893                     root_cl.launchThread();
894                 } catch (Exception ex) {
895                     System.out.println(ex.getMessage());
896                     throw new PropRequestFilterException(ex.getMessage());
897                 }
898         }
899
900         manager.setFilter(this);
901     }
902
903 }//PrivacyProtectionFilter.java
904
```

```
package org.w3c.www.protocol.http.privacy;

import java.util.*;
import java.io.*; //Uniquement pour les tests

/**
*   Cette classe représente l'ensemble des domaines contenus dans les Cookies
*   appartenant à la "liste blanche " et à la "liste noire".<p>
*
*   Il s'agit de deux 'Vector' qui contiennent les noms de domaine (champ "domain")
*   des Cookies acceptés ou refusés par le proxy.
*   <br>Ces vecteurs sont triés par ordre lexicographique.
*/

public class CookieLists {

    Vector list_white;
    Vector list_black;
    String cookielistsfilename;
    MAJCookieLists majC;

    public CookieLists() {
        this.list_white = new Vector();
        this.list_black = new Vector();
        this.majC = null;
        this.cookielistsfilename = "";
    }

    public void launchThread() {

        this.majC = new MAJCookieLists(this);

        try
        {
            majC.start();
        }
        catch (Exception e)
        {
            e.printStackTrace();
            System.out.println("Problème dans le lancement du thread de mis à jour"
                        +"de la liste des cookies refusés ou acceptables");
        }

    }

    /**
    *   <br>On va chercher les infos nécessaires (les champs "domain") de tous les cookies
    *   de la liste noire et de la liste blanche dans le fichier "Cookiesfile.txt".<br>
    *   Une exception est générée si le format du fichier "Cookiesfile" n'est pas conforme
    *   ou s'il n'a pas été trouvé.
    */
    public synchronized void loadCookieLists() throws Exception {

        //On récupère le String contenu dans le fichier "cookielistsfilename".
        String fileTxt = StringFromFile.getStringFromFile(cookielistsfilename);

        //On parse ce String et on met à jour les valeurs de la classe CookieLists.java.
        StringTokenizer st = new StringTokenizer(fileTxt);

        String errParsing = "Error at file \""+cookielistsfilename+"\" parsing";

        String token = "";
        String domain = "";
        boolean isComment = false; //Un commentaire est une suite de mots commençant
                                   //par "[" et se terminant par "]"

    synchronized (list_white)
    {
    synchronized (list_black)
    {
        if (list_white.size() != 0)
        {
            this.list_white.removeAllElements();
        }
        if (list_black.size() != 0)
        {
```

```
77              this.list_black.removeAllElements();
78          }
79
80          while (st.hasMoreTokens()) {
81
82              try {token = st.nextToken(); } catch (Exception e) { }
83
84              if  (token.toUpperCase().equals("")) {}
85              else {
86                      if (token.toUpperCase().equals("["))    {isComment = true;}
87                      else
88                      {
89                          if (token.toUpperCase().equals("]"))    {isComment = false;}
90                          else
91                          {
92                              if (! isComment)
93                              {
94                                  if (token.toUpperCase().equals("+"))
95                                  {
96                                      try
97                                      {
98                                          domain = st.nextToken();
99                                      }
100                                     catch (Exception e) {throw new Exception (errParsing);
101                                     //Rappel : on trie les champs 'domain' des Cookies par
102                                     //ordre lexicographique, dans la liste blanche.
103                                      if  (list_white.size()==0) {list_white.addElement(domain)
104                                     else
105                                     {
106                                         int i=0;
107             while ((i<list_white.size())&&(domain.compareTo((String)list_white.get(i))>0))
108                                         {
109                                             i++;
110                                         }//endWhile
111                                         list_white.insertElementAt(domain,i);
112                                     }//endElse
113                                 }
114
115                                 if (token.toUpperCase().equals("-"))
116                                 {
117                                     try
118                                     {
119                                         domain = st.nextToken();
120                                     }
121                                     catch (Exception e) {throw new Exception (errParsing);
122                                     //Rappel : on trie les champs 'domain' des Cookies par
123                                     //ordre lexicographique, dans la liste noire.
124                                 if  (list_black.size()==0)  {list_black.addElement(domain)
125                                     else
126                                     {
127                                         int i=0;
128             while ((i<list_black.size())&&(domain.compareTo((String)list_black.get(i))>0))
129                                         {
130                                             i++;
131                                         }//endWhile
132                                         list_black.insertElementAt(domain,i);
133                                     }//endElse
134                                 }
135                             }//endIf
136                         }//endElse
137                 }//endElse
138             }//endElse
139         }//endwhile()
140     }//synchronized
141     }//synchronized
142
143     }//loadCookieLists()
144
145     public boolean isEmpty(String type) {
146
147         if (type.equals("+"))
148         {
149             synchronized (list_white)
150             {
151                 if (list_white.size() == 0)
152                 {
```

```
                    return true;
                } else { return false; }
            }
        }

        if (type.equals("-"))
        {
            synchronized (list_black)
            {
                if (list_black.size() == 0)
                {
                    return true;
                } else { return false; }
            }
        }

        return true;

    }

    /**
    *   Cette méthode renvoie <i>true</i> si le champ 'domain' donné en paramètre
    *   (<i>domaine</i>) se trouve dans la liste noire <i>list_black</i> si le
    *   type est "+", ou dans la liste blanche <i>list_white</i> si le type est "-".
    *   <br>Elle renvoie <i>false</i> si aucun des champs 'domain' des Cookies dans la
    *   liste noire (si type = "-") ou dans la liste blanche (si type = "+") n'est
    *   identique à l'argument.
    */
    public boolean matchCookieLists(String domaine, String type) {

        String domain = domaine.toLowerCase();

        if (type.equals("+"))
        {
            int i=0;
            synchronized (list_white)
            {
                while (i<list_white.size())
                {
                    if (domain.indexOf(((String)list_white.get(i)).toLowerCase()) != -1)
                    {
                        return true;
                    }
                    i++;
                }//endWhile
            }
            return false;
        }

        if (type.equals("-"))
        {
            int i=0;
            synchronized (list_black)
            {
                while (i<list_black.size())
                {
                    if (domain.indexOf(((String)list_black.get(i)).toLowerCase()) != -1)
                    {
                        return true;
                    }
                    i++;
                }//endWhile
            }
            return false;
        }

        return false;

    }//matchCookieLists
```

```
229     /**
230     *   Cette méthode s'occupe d'afficher sur le terminal utilisateur la liste des
231     *   champs 'domain' des Cookies qui sont acceptés et refusés par le proxy Jigsaw.
232     */
233     public void afficherCookieLists() {
234
235         System.out.println("Taille de la liste des cookies acceptés : "+list_white.size())
236         for (int i = 0;i<list_white.size(); i++)
237         {
238             System.out.println("domaine : "+((String)list_white.get(i)));
239         }//endFor
240
241         System.out.println("Taille de la liste des cookies refusés : "+list_black.size());
242         for (int i = 0;i<list_black.size(); i++)
243         {
244             System.out.println("domaine : "+((String)list_black.get(i)));
245         }//endFor
246
247     }//afficherFichiers()
248
249 }//class CookieLists
250
251 //-----------------------------------------------------------------------------
252
253 class MAJCookieLists extends Thread {
254
255     private CookieLists hl;
256     boolean halt;
257
258     MAJCookieLists(CookieLists hl) {
259         this.hl = hl;
260         this.halt = false;
261         this.setDaemon(true);
262         this.setPriority(3);
263     }
264
265     public void run() {
266
267         while (! halt)
268         {
269             try
270             {
271                 this.sleep(60100l);//600001
272             }
273             catch (Exception e) {   e.printStackTrace();
274                                     System.out.println("Error at sleeping..."); }
275
276             try
277             {
278                 hl.loadCookieLists();
279             }
280             catch (Exception ex)
281             {
282                 System.out.println(ex.getMessage());
283                 ex.printStackTrace();
284             }
285         }
286     }//run()
287
288 }// class MAJCookieLists
```

```
package org.w3c.www.protocol.http.privacy;

import java.net.URL;

import java.util.*;
import java.io.*; //Uniquement pour les tests

/**
*   Cette classe représente l'ensemble des URLs (ou plutôt 'parties' d'URL)
*   appartenant à la liste noire.<p>
*
*   Il s'agit d'un 'Vector' qui contient les (ensembles d') URLs bloquées par le proxy.
*   <br>Ce vecteur est trié par ordre lexicographique.
*/

public class BlackList {

    Vector list;
    String blacklistfilename;
    MAJBlackList majB;

    public BlackList() {
        this.list = new Vector();
        this.majB = null;
        this.blacklistfilename = "";
    }

    public void launchThread() {

        this.majB = new MAJBlackList(this);

        try
        {
            majB.start();
        }
        catch (Exception e)
        {
            e.printStackTrace();
            System.out.println("Problème dans le lancement du thread de mis à jour"
                                                        +"de la liste noire");
        }

    }

    /**
    *   <br>On va chercher les infos ('parties d'URLs') contenues dans la liste noire
    *   à partir du fichier 'blacklistfile.txt'.<br>
    *   Une exception est générée si le format du fichier "blacklistfile" n'est pas
    *   conforme à la syntaxe ou s'il n'a pas été trouvé.
    */
    public synchronized void loadBlackList() throws Exception {

        // On récupère le String contenu dans le fichier "blacklistfilename".
        String fileTxt = StringFromFile.getStringFromFile(blacklistfilename);

        // On parse ce String et on met à jour les valeurs de la classe BlackList.java.
        StringTokenizer st = new StringTokenizer(fileTxt);

        String token = "";
        String partUrl = "";
        boolean isComment = false; // Un commentaire est une suite de mots commençant
                                   //par "[" et se terminant par "]"

    synchronized (list)
    {
        if (list.size() != 0)
        {
            this.list.removeAllElements();
        }

        while (st.hasMoreTokens()) {

            try {token = st.nextToken(); } catch (Exception e) { }

            if  (token.toUpperCase().equals("")) {}
            else {
```

```
                    if (token.toUpperCase().equals("["))    {isComment = true;}
                    else
                    {
                        if (token.toUpperCase().equals("]"))    {isComment = false;}
                        else
                        {
                            if (! isComment)
                            {
                                partUrl = token;
                                // Rappel : on trie les 'parties' d'URL
                                // par ordre lexicographique.
                                if  (list.size()==0)    {list.addElement(partUrl);}
                                else
                                {
                                    int i=0;
                                   while
                                   ((i<list.size())&&(partUrl.compareTo((String)list.get
                                   (i))>0))
                                    {
                                        i++;
                                    }//endWhile
                                    list.insertElementAt(partUrl,i);
                                }//endElse
                            }//endIf
                        }//endElse
                }//endElse
            }//endElse
        }//endwhile()
    }//synchronized

    }//loadBlackList()

    public boolean isEmpty() {

        synchronized (list)
        {
            if (list.size() == 0)
            {
                return true;
            } else { return false; }
        }

    }

    /**
    *   Cette méthode renvoie <i>true</i> si l'URL donnée en paramètre (<i>url</i>)
    *   appartient à un des ensembles d'URLS se trouvant dans la liste
    *   noire <i>list</i>.
    *   <br>Elle renvoie <i>false</i> si aucun des ensembles d'URLs de la liste noire ne
    *   correspond à l'argument.
    */
    public boolean matchBlackList(URL url) {

        String repUrl = url.toString().toLowerCase();
        int i=0;
        synchronized (list)
        {
            while (i<list.size())
            {
                if (repUrl.indexOf(((String)list.get(i)).toLowerCase()) != -1)
                {
                    return true;
                }
                i++;
            }//endWhile
        }
        return false;

    }//matchBlackList
```

```
    /**
    *   Cette méthode s'occupe d'afficher sur le terminal utilisateur la liste des
    *   ensembles d'URLs qui sont bloqués par le proxy Jigsaw.
    */
    public void afficherBlackList() {

        System.out.println("Taille de la liste noire : "+list.size());
        for (int i = 0;i<list.size(); i++)
        {
            System.out.println("ensemble d'URLs : "+((String)list.get(i)));
        }//endFor

    }//afficherFichiers()

}//class BlackList

//---------------------------------------------------------------------------

class MAJBlackList extends Thread {

    private BlackList hl;
    boolean halt;

    MAJBlackList(BlackList hl) {
        this.hl = hl;
        this.halt = false;
        this.setDaemon(true);
        this.setPriority(3);
    }

    public void run() {

        while (! halt)
        {
            try
            {
                this.sleep(600001);//600001
            }
            catch (Exception e) {   e.printStackTrace();
                                    System.out.println("Error at sleeping..."); }

            try
            {
                hl.loadBlackList();
            }
            catch (Exception ex)
            {
                System.out.println(ex.getMessage());
                ex.printStackTrace();
            }
        }
    }//run()

}// class MAJBlackList
```

```
// DomainTree.java
// $Id: DomainTree.java,v 1.11 2000/08/16 21:38:04 ylafon Exp $
// (c) COPYRIGHT MIT and INRIA, 1996.
// Please first read the full copyright statement in file COPYRIGHT.html

package org.w3c.www.protocol.http.privacy;

import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.io.InputStreamReader;
import java.io.OutputStreamWriter;
import java.io.PrintStream;
import java.io.Writer;

import java.util.Date;
import java.util.Enumeration;
import java.util.Hashtable;
import java.util.Vector;

import java.net.URL;

import org.w3c.www.protocol.http.HttpException;
import org.w3c.www.protocol.http.Reply;
import org.w3c.www.protocol.http.Request;

import org.w3c.www.http.HttpCookie;
import org.w3c.www.http.HttpCookieList;
import org.w3c.www.http.HttpFactory;
import org.w3c.www.http.HttpMessage;
import org.w3c.www.http.HttpRequestMessage;
import org.w3c.www.http.HttpSetCookie;
import org.w3c.www.http.HttpSetCookieList;

class DomainNode {

    Hashtable nodes       = null; // < String , DomainNode >
    HttpSetCookie cookies [] = null;
    int nbcookies         = 0;

    protected boolean sameCookies(HttpSetCookie c1, HttpSetCookie c2) {
    if (c2.getPath() == null && c1.getPath() == null)
        return (c1.getName().equals(c2.getName()));
    else  if (c2.getPath() == null || c1.getPath() == null)
        return false;
    else return ((c1.getName().equals(c2.getName())) &&
             (c1.getPath().equals(c2.getPath())));
    }

    protected void addCookie(HttpSetCookie cookie) {
    int i = 0;
    while (i < nbcookies) {
        if (sameCookies(cookie,cookies[i])) {
        cookies[i] = cookie;
        return;
        }
        i++;
    }
    if (nbcookies < cookies.length) {
        cookies[nbcookies++] = cookie;
    } else {
        HttpSetCookie ncookies [] = new HttpSetCookie[cookies.length +1];
        System.arraycopy(cookies,0,ncookies,0,cookies.length);
        ncookies[nbcookies++] = cookie;
        cookies = ncookies;
    }
    }

    protected void sync(FileWriter writer)
    throws IOException
    {
    if (nbcookies > 0) {
        for (int i = 0; i < nbcookies; i++) {
        if (cookies[i].getMaxAge() > 0)
```

```
 77              writer.write(DomainTree.cookie2String(cookies[i]));
 78          }
 79      }
 80      Enumeration e = nodes.elements();
 81      DomainNode node = null;
 82      while (e.hasMoreElements()) {
 83          node = (DomainNode)e.nextElement();
 84          node.sync(writer);
 85      }
 86      }
 87
 88      DomainNode() {
 89      nodes     = new Hashtable(2);
 90      cookies   = new HttpSetCookie[1];
 91      nbcookies = 0;
 92      }
 93
 94  }
 95
 96  class DomainTree {
 97
 98      Hashtable nodes        = null; // < String , DomainNode >
 99
100      protected static String cookie2String(HttpSetCookie cookie) {
101      Date date = new Date();
102      return (cookie.getDomain()+ //0
103          "\t"+
104          String.valueOf(cookie.getSecurity()).toUpperCase()+ //1
105          "\t"+
106          cookie.getPath()+ //2
107          "\t"+
108          (cookie.getMaxAge()*1000+date.getTime())+ //3
109          "\t"+
110          cookie.getVersion()+ //4
111          "\t"+
112          cookie.getName()+ //5
113          "\t"+
114          cookie.getValue()+ //6
115          "\n");
116      }
117
118      protected HttpSetCookie string2Cookie(String cookie[]) {
119      HttpSetCookie cook = new HttpSetCookie();
120      cook.setDomain(cookie[0]);
121      cook.setSecurity(Boolean.getBoolean(cookie[1].toLowerCase()));
122      cook.setPath(cookie[2]);
123      long expire = Long.parseLong(cookie[3]);
124      long now = (new Date()).getTime();
125      cook.setMaxAge((int) ((expire - now) / 1000));
126      cook.setVersion(Integer.parseInt(cookie[4]));
127      cook.setName(cookie[5]);
128      cook.setValue(cookie[6]);
129      return cook;
130      }
131
132      protected synchronized void loadCookies(File file)
133      throws FileNotFoundException
134      {
135      try {
136          FileReader reader = new FileReader(file);
137          String cookie[] = new String[8];
138          int i = 0;
139          int ch;
140          StringBuffer buffer = new StringBuffer(30);
141          while ((ch = reader.read()) != -1) {
142          switch (ch) {
143          case '#':
144              while ((ch = reader.read()) != '\n')
145              if (ch == -1) return;
146              break;
147          case '\n':
148              if (i > 0) {
149              cookie[i++] = buffer.toString();
150              HttpSetCookie setcookie = string2Cookie(cookie);
151              if (setcookie.getMaxAge() > 0)
152                  insertCookie(setcookie);
```

```
            buffer = new StringBuffer(30);
            cookie = new String[8];
            i = 0;
            }
            break;
        case '\t':
        case ' ':
            cookie[i++] = buffer.toString();
            buffer = new StringBuffer(30);
            break;
        default:
            buffer.append((char) ch);
        }
        }
    } catch (IOException ex) {
        System.out.println(ex.getMessage());
        ex.printStackTrace();
    }
    }

    protected synchronized void sync(File file) {
    Enumeration e     = nodes.elements();
    DomainNode node   = null;
    FileWriter writer = null;
    try {
        writer = new FileWriter(file);
        writer.write("# Jigsaw client HTTP Cookie File\n");
        writer.write("# This is a generated file!  Do not edit.\n\n");
        while (e.hasMoreElements()) {
        node = (DomainNode)e.nextElement();
        node.sync(writer);
        }
        writer.close();
    } catch (IOException ex) {
        System.out.println(ex.getMessage());
        ex.printStackTrace();
    } finally {
        try {if (writer != null) writer.close();} catch (Exception ex2) {}
    }
    }

    protected boolean isIp(String domain) {
    int last = domain.length()-1;
    return ((domain.charAt(last)>='0') && (domain.charAt(last)<='9'));
    }

    protected String[] domainParts(String dom) {
    if (dom == null)
        return null;
    String domain = new String(dom);
    Vector V = new Vector(5);
    int i = 0; int j = 0; int max = domain.length();
    // fix the . symbol bug
    while (i < max) {
        j = domain.indexOf('.', i);
        if (j == -1)
        j = max;
        V.addElement(domain.substring(i, j));
        i = j + 1;
    };
    // end of fix
    if (V.size() == 0)
        return null;
    String parts [] = new String[V.size()];
    V.copyInto(parts);
    return parts;
    }

    protected HttpCookie setCookie2Cookie(HttpSetCookie setcookie) {
    HttpCookie cookie = new HttpCookie();
    if (setcookie != null) {
        cookie.setName(setcookie.getName());
        cookie.setValue(setcookie.getValue());
        cookie.setDomain(setcookie.getDomain());
        cookie.setPath(setcookie.getPath());
        cookie.setVersion(setcookie.getVersion());
```

```
229         return cookie;
230     }
231     return null;
232     }
233
234     protected void addMatchingPathCookiesInVector(HttpSetCookie cookieArray[],
235                         String path,
236                         Vector V) {
237     int i = 0;
238     while ( i < cookieArray.length ) {
239         if (path.equals("/"))
240         V.addElement(setCookie2Cookie(cookieArray[i]));
241         else  if (cookieArray[i].getPath() != null) {
242         if (path.startsWith(cookieArray[i].getPath())) {
243             // transform SetCookie in Cookie
244             V.addElement(setCookie2Cookie(cookieArray[i]));
245         }
246         }
247         i++;
248     }
249     }
250
251     public HttpCookieList getCookies(URL url) {
252     String domain   = url.getHost();
253     String path     = url.getFile();
254     String parts [] = domainParts(domain);
255     if (parts == null) //FIXME Exception
256         return null;
257     Vector V = new Vector(5);
258     int i = 0;
259     DomainNode node = null;
260     Hashtable childs = nodes;
261
262     if (isIp(domain)) {
263         node = (DomainNode) childs.get(parts[i]);
264         while ( i < parts.length ) {
265         if (node == null)
266             return null;
267         if (node.nbcookies > 0) {
268             addMatchingPathCookiesInVector(node.cookies, path, V);
269         }
270         if ((i + 1) < parts.length)
271             node = (DomainNode)childs.get(parts[++i]);
272         else
273             node = null;
274
275         if (node == null)
276             break;
277         childs = node.nodes;
278         }
279     } else {
280         i = parts.length - 1;
281         node = (DomainNode) childs.get(parts[i]);
282         while ( i >= 0 ) {
283         if (node == null)
284             return null;
285         if (node.nbcookies > 0) {
286             addMatchingPathCookiesInVector(node.cookies, path, V);
287         }
288         if (i > 0)
289             node = (DomainNode)childs.get(parts[--i]);
290         else
291             node = null;
292         if (node == null)
293             break;
294         childs = node.nodes;
295         }
296     }
297     if (V.size() == 0)
298         return  null;
299     HttpCookie cookieArray[] = new HttpCookie[V.size()];
300     V.copyInto(cookieArray);
301     return HttpFactory.makeCookieList(cookieArray);
302     }
303
304     public void insertCookie(HttpSetCookie cookie) {
```

```
    String domain  = cookie.getDomain();
    String parts[] = domainParts(domain);
    if (parts == null) //FIXME Exception
        return;
    int i = 0;
    DomainNode node = null;
    Hashtable childs = nodes;
    if (isIp(domain)) {
        node = (DomainNode) childs.get(parts[i]);
        while ( true ) {
        if (node == null) {
            node = new DomainNode();
            childs.put(parts[i], node);
        }
        if (i == parts.length - 1) {
            node.addCookie(cookie);
            return;
        }
        node   = (DomainNode)childs.get(parts[++i]);
        if (node == null) {
            node = new DomainNode();
            childs.put(parts[i], node);
        }
        childs = node.nodes;
        }
    } else {
        i = parts.length - 1;
        node = (DomainNode) childs.get(parts[i]);
        while ( true ) {
        if (node == null) {
            node = new DomainNode();
            childs.put(parts[i], node);
        }
        if (i == 0) {
            node.addCookie(cookie);
            return;
        }
        node   = (DomainNode)childs.get(parts[--i]);
        if (node == null) {
            node = new DomainNode();
            childs.put(parts[i], node);
        }
        childs = node.nodes;
        }
    }
    }

    DomainTree() {
    this.nodes = new Hashtable(10);
    }
}
```